SPECIFICATION AND TOWERS IN SHIFT SPACES

VAUGHN CLIMENHAGA

Abstract. We show that shift spaces with a non-uniform specification property admit a tower with exponential tails for the unique equilibrium state. This gives strong statistical properties including the Bernoulli property, exponential decay of correlations, central limit theorem, and analyticity of pressure, which are new even for uniform specification. We give applications to shifts of quasi-finite type, synchronised shifts, and coded shifts. The proof goes via a structure theorem: a shift with non-uniform specification can be modeled by a strongly positive recurrent countable-state Markov shift to which every equilibrium state lifts.

Contents

Date: February 18, 2016.
2010 *Mathematics Subject Classification.* 37B10, 37D35 (primary), 37D25 (secondary).
The author is partially supported by NSF grant DMS-1362838.

1. Introduction

Consider a continuous map $f\colon X \to X$ on a compact metric space. An **equilibrium state** for a potential function $\varphi\colon X \to \mathbb{R}$ is an f-invariant measure maximising $h_\mu(f) + \int \varphi\, d\mu$.

Existence, uniqueness, and statistical properties of equilibrium states have consequences for many areas of dynamics and geometry, including SRB (physical) measures for smooth maps [Rue76]; entropy rigidity for geodesic flow [Kat82]; large deviations [Kif90]; distribution of geodesics in negative [Bow72] and non-positive curvature [Kni98]; multifractal analysis [BSS02]; the Weil–Petersson metric [McM08]; Teichmüller flow [BG11]; representation theory [BCLS15]; phase transitions and quasicrystals [BL13]; and diffusion along periodic surfaces [AHS14].

In this paper we consider the case where X is a shift space and f is the shift map σ. The most complete results are known when X is a mixing subshift of finite type: in this case every Hölder continuous potential has a unique equilibrium state, and this measure has strong statistical properties (Bernoulli property, exponential decay of correlations, central limit theorem) [Bow75]; moreover, the topological pressure function is analytic [PP90].

A weaker criterion for uniqueness is given by the **specification** condition [Bow74], which is a 'uniform mixing' property. Uniqueness results using non-uniform specification conditions have been proved by the author and D.J. Thompson [CT12, CT13], but the stronger statistical properties do not seem to have been studied using any version of specification. It is known that these stronger statistical properties hold for systems on which a certain 'tower' can be built [You98, You99]. Our main result is the following theorem, which uses a non-uniform specification condition to establish uniqueness and statistical properties by building a tower. Our conditions are given in terms of the **language** $\mathcal{L}$ of the shift X, which contains all finite words appearing in some element of X. See §2 for complete definitions.

Theorem 1.1. *Let X be a shift space on a finite alphabet and let $\varphi\colon X \to \mathbb{R}$ be Hölder. Suppose there are $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}(X)$ such that*

- **[I]** *there is $\tau \in \mathbb{N}$ s.t. for every $v, w \in \mathcal{G}$ there is $u \in \mathcal{L}_{\leq\tau}$ with $vuw \in \mathcal{G}$;*
- **[II]** *$P(\mathcal{C}^p \cup \mathcal{C}^s \cup (\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s), \varphi) < P(\varphi)$, where P is topological pressure;*
- **[III]** *there is $L \in \mathbb{N}$ such that if $u, v, w \in \mathcal{L}$ have $|v| \geq L$, $uvw \in \mathcal{L}$, $uv, vw \in \mathcal{G}$, then $v, uvw \in \mathcal{G}$.*

Then the following are true.

- (i) *(X, φ) has a unique equilibrium state μ.*
- (ii) *μ has the Gibbs property* (2.11) *with respect to φ and $\mathcal{G}$.*
- (iii) *μ is the limiting distribution of φ-weighted periodic orbits.*
- (iv) *Up to a finite period, (X, σ, μ) has the Bernoulli property and exponential decay of correlations for Hölder observables.*[1]

[1] If we replace "$\leq \tau$" with "$= \tau$" and assume that $\gcd\{|w| + \tau \mid w \in \mathcal{G}\} = 1$, then we can remove the words "up to a finite period"; see Theorem 3.2 and Remark 3.6.

(v) (X, σ, μ) *satisfies the central limit theorem for Hölder observables* ψ, *with variance 0 if and only if* ψ *is cohomologous to a constant.*
(vi) *Given any Hölder continuous* $\psi \colon X \to \mathbb{R}$, *there is* $\varepsilon > 0$ *such that the pressure function* $t \mapsto P(\varphi + t\psi)$ *is real analytic on* $(-\varepsilon, \varepsilon)$.

The classical specification property requires **[I]** to hold with $\mathcal{G} = \mathcal{L}$;[2] conclusions (i)–(iii) are well-known in this case [Bow74]. The **non-uniform specification** conditions **[I]**–**[III]** are mild variants of criteria given in [CT12, CT13] to extend these conclusions beyond the uniform setting. The idea is that $\mathcal{G}$ is a collection of **'good'** words for which specification holds, while $\mathcal{C}^p$ and $\mathcal{C}^s$ are collections of **prefixes** and **suffixes**; together with $\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s$ (the words with no **decomposition**), $\mathcal{C}^p$ and $\mathcal{C}^s$ give a list of **'obstructions to specification'**, which **[II]** requires to have small pressure.

Conclusions (iv)–(vi) are standard for mixing subshifts of finite type (SFTs) [Bow75], but do not appear to have been studied using any version of specification. For example, consider the **cocyclic subshifts** introduced by Kwapisz [Kwa00]; given square matrices $\Phi_1, \dots, \Phi_m$, the shift space $X = \{x \in \{1, \dots, m\}^{\mathbb{N}} \mid \Phi_{x_1} \cdots \Phi_{x_n} \neq \mathbf{0} \text{ for all } n\}$ is not always sofic, but has specification whenever it is transitive, and thus conclusions (i)–(vi) apply to (X, σ, φ) for every Hölder $\varphi \colon X \to \mathbb{R}$. This extends results from [Kwa04]. Similarly, shifts with specification were studied as examples of shifts of quasi-finite type by Buzzi [Buz05] and as synchronised shifts by Thomsen [Tho06]; Theorem 1.1 extends some of those results; see §3.2.1.

Theorem 1.1 is a consequence of the following structure theorem, which relates non-uniform specification to countable-state Markov shifts. Thermodynamic properties of these shifts are intimately related to **recurrence** properties of potential functions; we give more complete definitions in §2.1.3. The main fact we will need is that **strong positive recurrence** implies existence of a unique **Ruelle–Perron–Frobenius** (RPF) measure m on Σ, which has strong statistical properties [CS09] and can be interpreted as the unique equilibrium state for $\varphi \circ \pi$.

Theorem 1.2. *Let* X *be a shift space on a finite alphabet and let* $\varphi \colon X \to \mathbb{R}$ *be Hölder. If* (X, φ) *has the non-uniform specification conditions* **[I]**–**[III]**, *then there is a strongly irreducible countable-state Markov shift* Σ *and an injective 1-block code* $\pi \colon \Sigma \to X$ *such that*

(a) $\varphi \circ \pi$ *is strongly positive recurrent on* Σ; *and*
(b) *every equilibrium state* μ *for* (X, σ, φ) *has* $\mu = \pi_*\nu$ *for some shift-invariant* ν *on* Σ.

Combining (a) *and* (b), (X, σ, φ) *has a unique equilibrium state* μ, *and* $\mu = \pi_* m$ *where* m *is the RPF measure on* Σ *for* $\varphi \circ \pi$. *In particular, conclusions* (i)–(vi) *of Theorem 1.1 hold for* (X, φ). *The period in* (iv) *is given by the gcd of the lengths of periodic orbits in* Σ.

[2] Without the upper bound τ this is just transitivity. There are many versions of specification in the literature; this definition is specialised for the symbolic setting and is slightly weaker than Bowen's original one.

See §3.1 for a map of the proof of Theorem 1.2. The starting point is an argument of Bertrand for constructing a synchronising word using specification [Ber88]; although **[I]**–**[III]** do not imply that X is synchronised, we can still use **[I]** to 'synchronise good words' and produce a collection $\mathcal{F} \subset \mathcal{L}$ of words that can be **'freely concatenated'**, and use this collection to describe Σ. The lengthiest part of the proof is the argument that **[II]**–**[III]** can be used to get (a) and (b). Then conclusions (i)–(vi) follow from existing results in the literature.

One can also interpret the Markov shift Σ in Theorem 1.2 as a Young tower along the lines of [You99] by inducing on a single state of Σ. Then condition (a) on strong positive recurrence is equivalent to the condition that the tower have **'exponential tails'**, and (b) is the condition that every equilibrium state be **liftable** to the tower; see [PSZ14] for further discussion of this point of view. The effort we must expend in §§5–6 to prove (a) and (b) illustrates a general theme: *even when it is relatively clear how to build a tower, it is usually a non-trivial problem to verify that equilibrium states lift to the tower and that the tower's tails decay exponentially*; see [Kwa04, Buz05, Tho06] for symbolic examples, and [Kel89, PZ07, PSZ08, PS08, BT09, IT10] for smooth examples. One goal of the present approach is to give a set of more readily verifiable conditions that can establish liftability and exponential tails. This will be particularly valuable if it can be extended to the smooth setting; the non-uniform specification properties from [CT12, CT13] have been extended and applied in this setting, where along with a non-uniform expansivity property they once again yield uniqueness [CT14, CT15, CFT15]. Although the results given here are for symbolic systems, it is hoped that they will admit a similar generalisation; see §3.2.4.

The examples that motivated the introduction of non-uniform specification in [CT12] were subshift factors of β-shifts and S-gap shifts. Although condition **[III]** does not pass directly to factors, we can give an alternative set of conditions that do behave well. Theorem 1.3 below applies to subshift factors of β-shifts and S-gap shifts with $\varphi = 0$, giving a unique **measure of maximal entropy** (MME) satisfying (i)–(vi) (see §3.3.4 and Corollary 3.19); this extends [CT12, Theorem A], which only gave (i)–(iii). We also give applications to shifts of quasi-finite type and to synchronised shifts.

Given $\mathcal{C}^+, \mathcal{C}^- \subset \mathcal{L}$ and $M \in \mathbb{N}$, consider the collection[3]

$$\mathcal{G}(\mathcal{C}^\pm, M) := \{w \in \mathcal{L} \mid w_{[1,i]} \notin \mathcal{C}^-, w_{(|w|-i,|w|]} \notin \mathcal{C}^+ \ \forall M \le i \le |w|\} \tag{1.1}$$

of all words that do not start with a long element of $\mathcal{C}^-$ or end with a long element of $\mathcal{C}^+$. Say that $\mathcal{C}^\pm$ is a **complete list of obstructions to specification** if

[I*] for every $M \in \mathbb{N}$ there is $\tau = \tau(M)$ such that for all $v, w \in \mathcal{G}(\mathcal{C}^\pm, M)$ there is $u \in \mathcal{L}$ with $|u| \le \tau$ such that $vuw \in \mathcal{L}$.

[3] Note that this is not the collection $\mathcal{G}^M$ that appears in [CT12, CT13]; see §3.2.

Note that the word vuw need not be in $\mathcal{G}(\mathcal{C}^{\pm}, M)$ (cf. **[I]**), so enlarging $\mathcal{C}^{\pm}$ cannot cause **[I*]** to fail; in particular, every $\mathcal{C}^{\pm}$ that contains a complete list of obstructions is itself a complete list. Finally, we require $\mathcal{C}^{\pm}$ to satisfy the following condition:

$$(vw \in \mathcal{C}^+ \Rightarrow v \in \mathcal{C}^+) \quad \text{and} \quad (vw \in \mathcal{C}^- \Rightarrow w \in \mathcal{C}^-). \tag{1.2}$$

We will see that (1.2) holds in various natural classes of examples. The following result is proved in §7.

Theorem 1.3. *Let X be a shift space with language $\mathcal{L}$, and let $\varphi\colon X \to \mathbb{R}$ be Hölder. If $\mathcal{C}^{\pm} \subset \mathcal{L}$ is a complete list of obstructions to specification satisfying* (1.2) *and admitting the pressure bound $P(\mathcal{C}^- \cup \mathcal{C}^+, \varphi) < P(\varphi)$, then there are $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$ satisfying* **[I]**–**[III]**, *so (X, φ) satisfies the conclusions of Theorems 1.1 and 1.2.*

We apply Theorem 1.3 to shifts of quasi-finite type, which were introduced by Buzzi in [Buz05]. Say that $w \in \mathcal{L}$ is a **left constraint** if there is $v \in \mathcal{L}$ such that $w_{[2,|w|]}v \in \mathcal{L}$ but $wv \notin \mathcal{L}$; let $\mathcal{C}^\ell$ be the collection of left constraints. Let $\mathcal{C}^r$ be the collection of analogously defined **right constraints**. The shift space X is a **shift of quasi-finite type** (QFT) if at least one list of constraints has small entropy: $\min\{h(\mathcal{C}^\ell), h(\mathcal{C}^r)\} < h(\mathcal{L})$.

Topologically mixing QFTs may have multiple MMEs [Buz05, Lemma 4]. On the other hand, we will prove that the lists of constraints form a complete list of obstructions to specification, and thus Theorem 1.3 gives uniqueness if *both* lists of constraints have small entropy; the same conclusion is true for any QFT satisfying a stronger mixing condition.

Theorem 1.4. *Let X be a shift space on a finite alphabet and φ a Hölder potential on X.*

1. *If X is topologically transitive, then $\mathcal{C}^- = \mathcal{C}^r$ and $\mathcal{C}^+ = \mathcal{C}^\ell$ form a complete list of obstructions to specification and satisfy* (1.2)*. In particular, if $P(\mathcal{C}^\ell \cup \mathcal{C}^r, \varphi) < P(\varphi)$, then Theorem 1.3 applies.*
2. *Suppose $X^+ = \{x_1x_2\cdots \mid x \in X\}$ is* ***topologically exact****: for every $w \in \mathcal{L}$ there is $N \in \mathbb{N}$ such that $\sigma^N([w]) = X^+$. Then $\mathcal{C}^- = \emptyset$ and $\mathcal{C}^+ = \mathcal{C}^\ell$ form a complete list of obstructions to specification. In particular, if $P(\mathcal{C}^\ell, \varphi) < P(\varphi)$ then Theorem 1.3 applies.*

This result is proved in §3.3.1. Part (2) of Theorem 1.4 applies to many of the piecewise affine transformations studied by Buzzi in [Buz97]; these are maps $f\colon [0,1]^d \to [0,1]^d$ given by $f(x) = Ax + b \pmod{\mathbb{Z}^d}$, where $A\colon \mathbb{R}^d \to \mathbb{R}^d$ is an expanding linear map and $b \in \mathbb{R}^d$. Such maps admit natural symbolic codings; Buzzi showed that these codings are QFTs with $h(\mathcal{C}^\ell) < h(\mathcal{L})$, and that they are topologically exact if either (i) all eigenvalues of A exceed $1 + \sqrt{d}$ in absolute value, or (ii) A, b have all integer entries.[4] He proved that these examples have unique MMEs; Theorem 1.4(2) gives an

[4]Proposition 1 of [Buz97] only states that such f are topologically mixing, but the proof in [Buz97, §5.1, Lemma 5] gives topological exactness.

alternate proof of this result, and allows it to be extended to factors (see §3.3.4) and to a class of non-zero potential functions.

We consider two more classes of examples. A shift X is **synchronised** if there is $s \in \mathcal{L}$ such that $vs, sw \in \mathcal{L}$ implies $vsw \in \mathcal{L}$. Shifts with specification are synchronised [Ber88], but not vice versa. Synchronised shifts can have multiple equilibrium states even if they are topologically transitive. Theorem 1.1 gives the following simple criterion for uniqueness and statistical properties, which extends results of Thomsen [Tho06]; see §3.3.2 for a proof and discussion.

Theorem 1.5. *Let (X, σ) be a topologically transitive synchronised shift, and $s \in \mathcal{L}$ a synchronising word. Let $Y := \{x \in X \mid s \text{ does not appear in } x\}$. If $\varphi\colon X \to \mathbb{R}$ is Hölder and $P(Y, \varphi) < P(X, \varphi)$, then the conclusions of Theorems 1.1 and 1.2 hold for (X, σ, φ).*

The last class of examples we consider are the **coded shifts**, which includes all transitive synchronised shifts. Given a shift space X and a set of words $G \subset \mathcal{L}(X)$, let $G^\infty \subset X$ be the set of all infinite concatenations of elements of G; that is,

$$G^\infty = \{x \in X \mid \text{there is } (n_k)_{k\in\mathbb{Z}} \subset \mathbb{Z} \text{ such that } n_k < n_{k+1} \\ \text{and } x_{[n_k, n_{k+1})} \in G \text{ for all } k \in \mathbb{Z}.\}$$

Recall that X is **coded** if there is a **generating set** $G \subset \mathcal{L}(X)$ such that $X = \overline{G^\infty}$.[5] Say that G is **uniquely decipherable** if whenever $u^1 u^2 \cdots u^m = v^1 v^2 \cdots v^n$ with $u^i, v^j \in G$, we have $m = n$ and $u^j = v^j$ for all j [LM95, Definition 8.1.21].

Theorem 1.6. *Let X be a coded shift on a finite alphabet and φ a Hölder potential on X. If X has a uniquely decipherable generating set G such that $\mathcal{D} = \mathcal{D}(G) := \{w \in \mathcal{L} \mid w \text{ is a subword of some } g \in G\}$ satisfies $P(\mathcal{D}, \varphi) < P(\varphi)$, then (X, φ) satisfies the conclusions of Theorems 1.1 and 1.2, with one exception: the coding map π may not be injective, but it is still finite-to-one μ-a.e. for the unique equilibrium state μ.*

The condition of unique decipherability in Theorem 1.6 is related to **[III]**, but does not imply it, so Theorems 1.1 and 1.2 cannot be applied directly to get Theorem 1.6; see §3.3.3.

In §2 we give the definitions and background used in the hypotheses and conclusion of Theorem 1.1. In §3 we give some intermediate results, a discussion of the motivations for Theorem 1.1, more details of the applications (including proofs of Theorems 1.4 and 1.5), and some open problems that remain. §4 contains preparatory results for the main proofs, including mild strengthenings of the Birkhoff and Shannon–McMillan–Breiman theorems (Theorems 4.1 and 4.2) that hold quite generally, not just in the setting of

[5] Equivalently, X is the closure of a uniformly continuous image of a countable-state irreducible topological Markov chain [FF92].

this paper. The proof of Theorem 1.2 (and hence Theorem 1.1) is given in §§5–6; Theorem 1.6 is proved in §5. In §7 we collect the remaining proofs, including the proof of Theorem 1.3.

Acknowledgments. I am grateful to the anonymous referee for a very careful reading and for many comments that improved the exposition, as well as pointing out an error in an earlier version of the result on factors. I am also grateful to Omri Sarig for clarifying aspects of strong positive recurrence as they appear in §2.1.3, and to Dominik Kwietniak for introducing me to cocyclic subshifts and [Kwa00, Kwa04].

2. Definitions

2.1. **Shift spaces and thermodynamic formalism.**

2.1.1. *Definitions and notation from symbolic dynamics.* We recall basic definitions from symbolic dynamics; see [LM95] for further details. Let A be a finite set (the **alphabet**) and let $\sigma\colon A^{\mathbb{Z}} \to A^{\mathbb{Z}}$ be defined by $\sigma(x)_k = x_{k+1}$. Define $\sigma\colon A^{\mathbb{N}\cup\{0\}} \to A^{\mathbb{N}\cup\{0\}}$ similarly.[6] Equip $A^{\mathbb{Z}}$ and $A^{\mathbb{N}\cup\{0\}}$ with the metric $d(x,y) = e^{-n(x,y)}$, where $n(x,y) = \min\{|k| \mid x_k \neq y_k\}$.

Let X be a shift space on A; that is, a closed σ-invariant subset of $A^{\mathbb{Z}}$ or $A^{\mathbb{N}\cup\{0\}}$. A **word** is a finite sequence of symbols from A (we allow the empty word, which has no symbols). The **language** $\mathcal{L} = \mathcal{L}(X)$ is the set of words that appear in some $x \in X$, so $\mathcal{L} = \bigcup_{n\geq 0} \mathcal{L}_n$, where

$$\mathcal{L}_n = \mathcal{L}_n(X) = \{w \in A^n \mid \text{there is } x \in X \text{ such that } x_1 \cdots x_n = w\}.$$

We write $A^* = \bigcup_{n\geq 0} A^n$ for the collection of all finite concatenations of elements of A, so $\mathcal{L}(X) \subset A^*$, with equality if and only if X is the full shift. When we need to work with an indexed collection of words, we will write the indices as superscripts; thus w^1, w^2 represent two different words, while w_1, w_2 represent the first and second symbols in the word w.

Juxtaposition denotes concatenation and will be used liberally throughout the paper both for words and for collections of words; for example, given $w \in A^*$, we will have occasion to refer to the following sets, or similar ones:

$$w\mathcal{L} \cap \mathcal{L} = \text{the set of all words in } \mathcal{L} \text{ that begin with } w,$$
$$\mathcal{L} \setminus \mathcal{L}w\mathcal{L} = \text{the set of all words in } \mathcal{L} \text{ that do not contain } w \text{ as a subword.}$$

Given a word $w \in A^*$, we write $|w|$ for the length of w; given $1 \leq i \leq j \leq |w|$, we write $w_{[i,j]} = w_i \cdots w_j$. When convenient, we use the notation $w_{(i,j]} = w_{[i+1,j]}$, and similarly for $w_{[i,j)}$ and $w_{(i,j)}$. We will use the same notation for subwords of an infinite sequence $x \in X$. Given a collection $\mathcal{D} \subset \mathcal{L}$, we write

$$\mathcal{D}_n = \mathcal{D} \cap \mathcal{L}_n = \{w \in \mathcal{D} \mid |w| = n\}, \qquad \mathcal{D}_{\leq n} = \{w \in \mathcal{D} \mid |w| \leq n\}.$$

[6] All of our results apply to both one- and two-sided shift spaces.

2.1.2. *Thermodynamic formalism for shift spaces.* We recall some basic definitions from thermodynamic formalism (adapted to the symbolic setting) and describe the notation that we will use. See [Wal82, CT12, CT13] for further details.

Let X be a shift space on a finite alphabet and $\varphi\colon X \to \mathbb{R}$ a continuous function, called a **potential**. Given φ, we define a function $\hat{\varphi}\colon \mathcal{L} \to \mathbb{R}$ by

$$\hat{\varphi}(w) = \sup_{x \in [w]} S_{|w|}\varphi(x), \tag{2.1}$$

where $[w] = \{x \in X \mid x_{[0,|w|)} = w\}$ is the **cylinder** defined by w, and $S_n\varphi(x) = \sum_{j=0}^{n-1} \varphi(\sigma^j x)$ is the nth Birkhoff sum. Given a collection of words $\mathcal{D} \subset \mathcal{L}$, we write

$$\begin{aligned} \Lambda_n(\mathcal{D},\varphi) &= \sum_{w\in\mathcal{D}_n} e^{\hat{\varphi}(w)}, \\ P(\mathcal{D},\varphi) &= \overline{\lim_{n\to\infty}} \frac{1}{n} \log \Lambda_n(\mathcal{D},\varphi). \end{aligned} \tag{2.2}$$

The first quantity is the **partition sum** of φ on $\mathcal{D}$, and the second is the **pressure** of φ on $\mathcal{D}$. When $\mathcal{D} = \mathcal{L}$ this gives the standard definition of topological pressure, and we write $P(\varphi) = P(\mathcal{L},\varphi)$. When $\varphi = 0$ we write $h(\mathcal{D}) = P(\mathcal{D}, 0)$ for the **entropy** of $\mathcal{D}$.

We will frequently use the following consequence of (2.2):

$$P(\mathcal{C} \cup \mathcal{D}, \varphi) = \max\{P(\mathcal{C},\varphi), P(\mathcal{D},\varphi)\} \text{ for every } \mathcal{C},\mathcal{D} \subset \mathcal{L}. \tag{2.3}$$

The variational principle [Wal82, Theorem 9.10] states that $P(\varphi)$ is the supremum of the quantities $h(\mu) + \int \varphi \, d\mu$ taken over all σ-invariant Borel probability measures, where $h(\mu)$ is the measure-theoretic entropy. The supremum is achieved at **equilibrium states**. Because σ is expansive, equilibrium states exist for all continuous φ, and so the real questions are uniqueness and statistical properties. These require a regularity condition on φ. Given $\beta > 0$, consider the following set of Hölder continuous functions:

$$\begin{aligned} C_\beta(X) = \{\varphi\colon X \to \mathbb{R} \mid {} & \text{there exists } |\varphi|_\beta \in \mathbb{R} \text{ such that} \\ & |\varphi(x) - \varphi(y)| \leq |\varphi|_\beta \, e^{-\beta n} \text{ whenever } x_k = y_k \text{ for all } |k| \leq n\}. \end{aligned} \tag{2.4}$$

We write $C^h(X) = \bigcup_{\beta>0} C_\beta(X)$ for the set of all Hölder functions.

Uniqueness of the equilibrium state was shown in [Bow74] when $\varphi \in C^h(X)$ and (X,σ) has specification. Uniqueness in the case of non-uniform specification was shown in [CT13]. The purpose of this paper is to establish statistical properties in these settings.

2.1.3. *Countable-state Markov shifts.* Given a directed graph with a countably infinite vertex set V, one obtains a **countable-state Markov shift** $\Sigma \subset V^{\mathbb{Z}}$ by the condition that $\mathbf{z} \in V^{\mathbb{Z}}$ is in Σ iff the graph contains an edge from $\mathbf{z}_n$ to $\mathbf{z}_{n+1}$ for every $n \in \mathbb{Z}$. The shift is **strongly irreducible** if for any $a, b \in V$ there is a path in the graph that leads from a to b. We will use the thermodynamic formalism for these shifts developed by Sarig

in [Sar99, Sar01]. Note that although Sarig's results are formulated for one-sided shifts, the parts of them that we will need extend to two-sided shifts by standard techniques; see §5.2 and §5.5.

Given a countable-state Markov shift Σ and a shift X on a finite alphabet A, we will be interested in **1-block codes** $\pi\colon \Sigma \to X$; this means that there is a map $\tau\colon V \to A$ such that $\pi(\mathbf{z})_n = \tau(\mathbf{z}_n)$ for every $n \in \mathbb{Z}$.

As in §2.1.2, we write $C_\beta(\Sigma)$ for the set of all functions $\Phi\colon \Sigma \to \mathbb{R}$ such that there is $|\Phi|_\beta > 0$ that makes the following hold for every $n \geq 0$.

$$|\Phi(\mathbf{z}) - \Phi(\mathbf{z}')| \leq |\Phi|_\beta \, e^{-\beta n} \text{ whenever } \mathbf{z}_k = \mathbf{z}'_k \text{ for all } |k| \leq n. \tag{2.5}$$

If (2.5) holds for all $n \geq 1$, but not necessarily for $n = 0$, we say that Φ is **locally Hölder continuous** [Sar99]. Given $\Phi \in C^h(\Sigma) := \bigcup_{\beta>0} C_\beta(\Sigma)$ and $a \in V$, we follow [Sar99] and write

$$Z_n(\Phi, a) := \sum_{T^n\mathbf{z}=\mathbf{z};\ \mathbf{z}_0=a} e^{S_n\Phi(\mathbf{z})}, \tag{2.6}$$

$$Z_n^*(\Phi, a) := \sum_{\substack{T^n\mathbf{z}=\mathbf{z};\ \mathbf{z}_0=a \\ \mathbf{z}_1,\dots,\mathbf{z}_{n-1}\neq a}} e^{S_n\Phi(\mathbf{z})}. \tag{2.7}$$

The **Gurevich pressure** of Φ is

$$P_G(\Phi) := \lim_{n\to\infty} \frac{1}{n} \log Z_n(\Phi, a); \tag{2.8}$$

this limit is independent of a. The potential Φ is **positive recurrent** if there is $C > 0$ with[7]

$$C^{-1} e^{nP(\varphi)} \leq Z_n(\Phi, a) \leq C e^{nP(\varphi)} \text{ for all } n \in \mathbb{N}, \tag{2.9}$$

and **strongly positive recurrent** if

$$\varlimsup_{n\to\infty} \frac{1}{n} \log Z_n^*(\Phi, a) < P_G(\Phi). \tag{2.10}$$

The definition of strong positive recurrence in [Sar01] is given in terms of positivity of a certain **determinant** $\Delta_a[\Phi]$. Equivalence of the two definitions follows from [Sar01] but is not explicitly stated there; for completeness we prove in §7 that $\Delta_a[\Phi] > 0$ iff (2.10) holds.

2.2. **Statistical properties.**

We recall several strong statistical properties that a measure can have; Theorem 1.1 says that these are all satisfied for the unique equilibrium state produced by **[I]**–**[III]**.

[7] In [Sar01] the equivalent condition $\sum nZ_n^*(\Phi, a)e^{-nP_G(\Phi)} < \infty$ is used; we follow [Sar99, Definition 2].

2.2.1. *Gibbs property.* Given $\mathcal{G} \subset \mathcal{L}$ and $\varphi\colon X \to \mathbb{R}$, we say that μ has the **Gibbs property for φ with respect to $\mathcal{G}$** if there is $Q_1 > 0$ such that[8]

$$\begin{aligned} &\text{for every } w \in \mathcal{L}, \text{ we have } \mu[w] \leq Q_1 e^{-|w|P(\varphi)+\hat\varphi(w)}; \text{ and} \\ &\text{for every } w \in \mathcal{G}, \text{ we have } \mu[w] \geq Q_1^{-1} e^{-|w|P(\varphi)+\hat\varphi(w)}. \end{aligned} \tag{2.11}$$

Note that the lower bound is only required to hold on $\mathcal{G}$, while the upper bound holds on all of $\mathcal{L}$. This weakened version of the classical Gibbs property was introduced in [CT12, CT13].

2.2.2. *Periodic orbits.* For $n \in \mathbb{N}$, let $\mathrm{Per}_n = \{x \in X \mid \sigma^n x = x\}$ be the set of n-periodic points. Note that Per_n is finite. Let

$$\mu_n = \frac{1}{\sum_{k=1}^n \sum_{x\in \mathrm{Per}_k} e^{S_k\varphi(x)}} \sum_{k=1}^n \sum_{x \in \mathrm{Per}_k} e^{S_k\varphi(x)} \delta_x \tag{2.12}$$

be the **φ-weighted periodic orbit measure** corresponding to periodic orbits of length at most n. Say that μ is the **limiting distribution of φ-weighted periodic orbits** if μ_n converges to μ in the weak* topology.

2.2.3. *Bernoulli property.* Given a state space S and a probability vector $p = (p_a)_{a\in S}$, the **Bernoulli scheme** with probability vector p is $(S^{\mathbb{Z}}, \sigma, \mu_p)$, where σ is the left shift map and $\mu_p[w] = \prod_{i=1}^{|w|} p_{a_i}$ for every $w \in S^*$. The measure μ is said to have the **Bernoulli property** if (X, σ, μ) is measure-theoretically isomorphic to a Bernoulli scheme.

We say that (X, σ, μ) has the Bernoulli property **up to a period** if there are $d \in \mathbb{N}$ and a subshift $Y \subset X$ such that

$$\sigma^d(Y) = Y \text{ and } \mu(\sigma^j(Y) \cap Y) = 0 \text{ for all } j = 1, 2, \dots, d-1, \tag{2.13}$$

and moreover (Y, σ^d, ν) has the Bernoulli property, where $\nu = \frac{1}{\mu(Y)}\mu|_Y = \mu|_Y \cdot d$. In particular, this means that (X, σ, μ) is measure-theoretically isomorphic to the direct product of a Bernoulli scheme and a finite rotation.

2.2.4. *Decay of correlations.* Let (X, σ, μ) be a shift space with an invariant measure μ. Given $\psi_1, \psi_2\colon X \to \mathbb{R}$, consider the **correlation functions**

$$\mathrm{Cor}_n^\mu(\psi_1, \psi_2) = \int (\psi_1 \circ \sigma^n)\psi_2 \, d\mu - \int \psi_1 \, d\mu \int \psi_2 \, d\mu.$$

We say that the system has **exponential decay of correlations** for observables in $C_\beta(X)$ if there is $\theta \in (0,1)$ such that for every $\psi_1, \psi_2 \in C_\beta(X)$ there is $K(\psi_1, \psi_2) > 0$ such that

$$|\mathrm{Cor}_n^\mu(\psi_1, \psi_2)| \leq K(\psi_1, \psi_2)\theta^{|n|} \text{ for every } n \in \mathbb{Z}. \tag{2.14}$$

As with the Bernoulli property, we say that (X, σ, μ) has exponential decay of correlations **up to a period** if there are $d \in \mathbb{N}$ and a subshift $Y \subset X$

[8] We will use $Q_1, Q_2, \dots$ to denote 'global constants' that are referred to throughout the paper. We will use K or C for 'local constants' that appear only within the proof of a given lemma or proposition, and are not used for more than one or two paragraphs.

satisfying (2.13) such that (Y, σ^d, ν) has exponential decay of correlations, where once again $\nu = \frac{1}{\mu(Y)}\mu|_Y = \mu|_Y \cdot d$.

2.2.5. *Central limit theorem.* Given (X, σ, μ) as above and $\psi \colon X \to \mathbb{R}$, we say that the **central limit theorem** holds for ψ if $\frac{1}{\sqrt{n}} S_n(\psi - \int \psi \, d\mu)$ converges in distribution to a normal distribution $\mathcal{N}(0, \sigma_\psi)$ for some $\sigma_\psi \geq 0$; that is, if

$$\lim_{n\to\infty} \mu\left\{x \mid \frac{1}{\sqrt{n}}\left(S_n\psi(x) - n\int \psi \, d\mu\right) \leq \tau\right\} = \frac{1}{\sigma_\psi\sqrt{2\pi}} \int_{-\infty}^{\tau} e^{-t^2/(2\sigma_\psi^2)} \, dt$$

for every $\tau \in \mathbb{R}$. (When $\sigma_\psi = 0$ the convergence is to the Heaviside function.)

Say that ψ is **cohomologous to a constant** if there are a measurable function $u \colon X \to \mathbb{R}$ and a constant $c \in \mathbb{R}$ such that $\psi(x) = u(x) - u(\sigma x) + c$ for μ-a.e. $x \in X$. One generally expects that in the central limit theorem the variance σ_ψ^2 is 0 if and only if ψ is cohomologous to a constant. This will hold in our setting as well.[9]

3. Description of approach, further results, and discussion

3.1. **Intermediate results.** We will prove Theorem 1.2 (and hence Theorem 1.1) via Theorems 3.1 and 3.2 below. The strategy is to use [**I**]–[**III**] to produce $\mathcal{F} \subset \mathcal{L}$ satisfying the following **free concatenation** property, which strengthens [**I**].

[$\mathbf{I}_0$] Given any $v, w \in \mathcal{F}$ we have $vw \in \mathcal{F}$.

Once the collection $\mathcal{F}$ has been produced there is a natural way to construct a countable-state Markov shift Σ that embeds into X. Writing

$$I = I(\mathcal{F}) := \mathcal{F} \setminus \mathcal{F}\mathcal{F} = \{w \in \mathcal{F} \mid w \neq uv \text{ for any non-trivial } u, v \in \mathcal{F}\} \tag{3.1}$$

for the set of **irreducible** elements of $\mathcal{F}$,[10] we use the alphabet

$$A_I = \{(w, k) \in I \times \mathbb{N} \mid w \in I \text{ and } 1 \leq k \leq |w|\}, \tag{3.2}$$

and think of (w, k) as representing the state "we are currently in the word w, and have seen the first k symbols of w". Let $\Sigma = \Sigma(\mathcal{F}) \subset (A_I)^{\mathbb{Z}}$ be the Markov shift given by allowing the following transitions:

$$\begin{aligned} (w, k) &\to (w, k+1) \text{ for any } w \in I \text{ and } 1 \leq k < |w|, \\ (w, |w|) &\to (v, 1) \text{ for any } w, v \in I. \end{aligned} \tag{3.3}$$

Define a one-block code $\pi \colon \Sigma \to X$ by $(w, k) \mapsto w_k$, the kth symbol of w.

The above construction can also be given starting with I instead of $\mathcal{F}$; if $I \subset \mathcal{L}$ is such that $I^* \subset \mathcal{L}$ and $I \cap II = \emptyset$, then one can recover $\mathcal{F}$ as I^* and construct (Σ, π) as above.

[9]One could likely go further and derive a **Green–Kubo formula** expressing the variance σ_ψ as a sum of correlations [Liv96, Theorem 1.1(2)], or deduce an **almost sure invariance principle** [MN05], but since the main work in this paper is the construction of the tower itself, we will not discuss these here.

[10]We adopt the convention that $\mathcal{F}$ does not contain the empty word.

Theorem 3.1 below clarifies the relationship between properties of $\mathcal{F}$ and properties of (Σ, π); in particular, it gives sufficient conditions for conclusions (a) and (b) of Theorem 1.2. Theorem 3.2 uses **[I]**–**[III]** to produce $\mathcal{F}$ satisfying these conditions. To state the theorems, we need the following strengthening of **[II]**, given in terms of I and $\mathcal{F} = I^*$.

[II′] $P(I, \varphi) < P(\varphi)$, and there are $\mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L}$ with $P(\mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p\mathcal{F}\mathcal{E}^s), \varphi) < P(\varphi)$.

We also need the following variant of **[III]**.

[III*] If $x \in X$ and $i \leq j \leq k \leq \ell$ are such that $x_{[i,k)}, x_{[j,\ell)} \in \mathcal{F}$, and there are $a < j$ and $b > k$ such that $x_{[a,j)}, x_{[k,b)} \in \mathcal{F}$, then $x_{[j,k)} \in \mathcal{F}$.

To compare **[III*]** and **[III]** it is helpful to reformulate the latter as follows.

[III] There is $L \in \mathbb{N}$ such that if $x \in X$ and $i \leq j \leq k \leq \ell \in \mathbb{Z}$ are such that $k - j \geq L$ and $x_{[i,k)}, x_{[j,\ell)} \in \mathcal{G}$, then $x_{[j,k)}, x_{[i,\ell)} \in \mathcal{G}$.

The naturality of **[III]** and **[III*]** is discussed in §3.2.4, and the question of their necessity is discussed in §3.3.3. Both conditions are illustrated in Figure 3.1. Note that although i, j, k, ℓ must appear in the order shown, a can be to either side of i (or equal to it), so long as $a < j$, and similarly there is no constraint on b and ℓ. Moreover, there is no requirement in **[III*]** that $k - j$ be large; in practice we will produce $\mathcal{F}$ such that every overlap has length at least L, so the case $k - j < L$ will never arise.

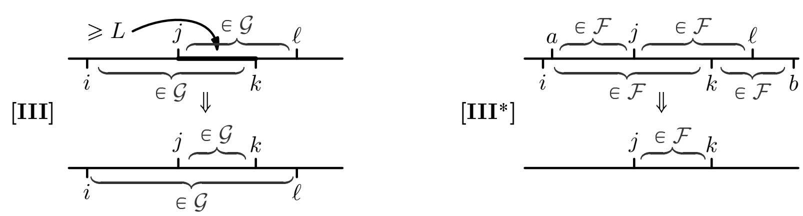


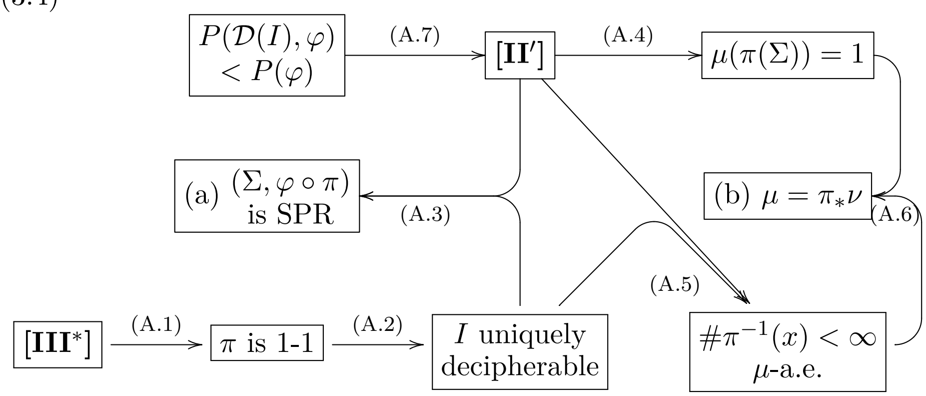
FIGURE 3.1. Conditions **[III]** and **[III*]**.

Beyond **[II′]** and **[III*]**, there are several more conditions and conclusions that enter the statement of Theorem 3.1.

- Injectivity of π (equivalent to the tower's induced map being a first return map).
- Unique decipherability: $w \in \mathcal{F}$ has only one representation in I^*.
- Strong positive recurrence (SPR) of $(\Sigma, \varphi \circ \pi)$; this is conclusion (a) of Theorem 1.2.
- Charging the tower: every equilibrium state μ has $\mu(\pi(\Sigma)) = 1$.
- Multiplicity bound: $\pi^{-1}(x)$ is finite μ-a.e.
- Liftability: every equilibrium state μ for (X, φ) has $\mu = \pi_*\nu$ for some shift-invariant ν on Σ; this is conclusion (b) of Theorem 1.2.
- A pressure gap $P(\mathcal{D}, \varphi) < P(\varphi)$ for $\mathcal{D} = \mathcal{D}(I) := \{w \in \mathcal{L} \mid uwv \in I$ for some $u, v \in \mathcal{L}\}$.

The relationship between these is illustrated by the following graph, and formalised by Theorem 3.1(A) below. The properties in the last column of (3.4) are to hold for every equilibrium state μ. The labels on the arrows indicate which part of Theorem 3.1 establishes the corresponding implication.

(3.4)



The following result establishes Theorem 1.6, and gives one half of the proof of Theorems 1.1 and 1.2.

Theorem 3.1. *Let X be a shift space on a finite alphabet, and $\varphi \in C^h(X)$.*

(A) *Suppose that $\mathcal{F} \subset \mathcal{L}(X)$ satisfies $[\mathbf{I}_0]$ and let $I = \mathcal{F} \setminus \mathcal{F}\mathcal{F}$; equivalently, consider $I \subset \mathcal{L}$ with the property that $II \cap I = \emptyset$ and $\mathcal{F} := I^* \subset \mathcal{L}$. Let Σ, π be as above.*
  1. *If $\mathcal{F}$ satisfies $[\mathbf{III}^*]$ then π is 1-1.*
  2. *If π is 1-1 then I is uniquely decipherable.*
  3. *If I is uniquely decipherable and $[\mathbf{II}']$ holds, then $(\Sigma, \varphi \circ \pi)$ is strong positive recurrent, giving conclusion* (a) *of Theorem 1.2.*
  4. *If $[\mathbf{II}']$ holds, then every equilibrium state μ for (X, φ) has $\mu(\pi(\Sigma)) = 1$.*[11]
  5. *If I is uniquely decipherable and $[\mathbf{II}']$ holds, then every equilibrium state μ for (X, φ) has $\#\pi^{-1}(x) < \infty$ for μ-a.e. $x \in X$.*
  6. *If μ is a σ-invariant probability measure on $\pi(\Sigma)$ such that $\#\pi^{-1}(x) < \infty$ for μ-a.e. x, then there is a shift-invariant probability measure ν on Σ such that $\mu = \pi_* \nu$. In particular, conclusion* (b) *of Theorem 1.2 holds.*
  7. *If I generates $\mathcal{L}$ and $P(\mathcal{D}(I), \varphi) < P(\varphi)$, then $[\mathbf{II}']$ holds.*

(B) *If $\mathcal{F}$ satisfies $[\mathbf{I}_0], [\mathbf{II}'], [\mathbf{III}^*]$, or if I is a uniquely decipherable generating set for $\mathcal{L}$ with $P(\mathcal{D}(I), \varphi) < P(\varphi)$, then by part* (A), *$(\Sigma, \pi)$ satisfies conclusions* (a) *and* (b) *of Theorem 1.2; moreover, conclusions* (i)–(vi) *of Theorem 1.1 hold for (X, φ), where the Gibbs property in* (ii) *is w.r.t. $\mathcal{F}$, and the period in* (iv) *is given by* $\gcd\{|w| \mid w \in \mathcal{F}\}$.

[11]If π is 1-1, this immediately implies the liftability condition $\mu = \pi_* \nu$ by taking $\nu := (\pi^{-1})_* \mu$. Without injectivity the situation is more delicate.

The other half of the proofs of Theorems 1.1 and 1.2 is given by Theorem 3.2 below, which uses a slightly weaker version of **[III]** that is useful in some applications (see [CP16]).

[III$_\mathrm{a}$] There is L s.t. if $uv, vw \in \mathcal{G}$, $|v| \geq L$, and $uvw \in \mathcal{L}$, then $v \in \mathcal{G}$.
[III$_\mathrm{b}$] There is L such that if $uv, vw \in \mathcal{G}$, $|v| \geq L$, and $xuvw \in \mathcal{G}$ for some $x \in \mathcal{L}$, then $uvw \in \mathcal{G}$.

Note that **[III$_\mathrm{a}$]** and **[III$_\mathrm{b}$]** both follow from **[III]**, but do not necessarily imply it. In fact we will use **[III$_\mathrm{a}$]** repeatedly throughout the proof, while **[III$_\mathrm{b}$]** only appears once (see the footnote at the end of §6.2.2).

Theorem 3.2. *Let X be a shift space on a finite alphabet and let $\varphi\colon X \to \mathbb{R}$ be Hölder continuous. Suppose there are $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}(X)$ satisfying* **[I]**, **[II]**, **[III$_\mathrm{a}$]**, *and* **[III$_\mathrm{b}$]**. *Then there are $\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s \subset \mathcal{L}$ satisfying* **[I$_0$]**, **[II$'$]**, *and* **[III*]** *such that a measure μ has the Gibbs property for $\mathcal{G}$ if and only if it has the Gibbs property for $\mathcal{F}$. If in addition $\mathcal{G}$ satisfies*

[I$'$] *there is $\tau \in \mathbb{N}$ s.t. for all $v, w \in \mathcal{G}$ there is $u \in \mathcal{L}_\tau$ with $vuw \in \mathcal{G}$,*

then $\gcd\{|w| \mid w \in \mathcal{F}\} = \gcd\{|v| + \tau \mid v \in \mathcal{G}\}$*; see Remark 3.6.*

Remark 3.3. Although Theorem 3.2 guarantees that $\mathcal{E} := I \cup \mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p\mathcal{F}\mathcal{E}^s)$ has $P(\mathcal{E}, \varphi) < P(\varphi)$ (condition **[II$'$]**), we may have $P(\mathcal{E}, \varphi) > P(\mathcal{C}, \varphi)$. This happens already for SFTs; let X be the SFT on the alphabet $\{1, \dots, k\}$ defined by allowing a sequence x if and only if $x_{n+1} - x_n = 1$ or $2 \pmod k$ for every n. Then $h(X) = \log 2$, and we show in §7 that if $\mathcal{F} \subset \mathcal{L}(X)$ is any collection satisfying **[I$_0$]**, then for every choice of $\mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L}$ we have $h(\mathcal{E}) \geq h(X) - \frac{4 \log 2}{k}$. This is despite the fact that $\mathcal{L}$ itself has specification and so we can take $\mathcal{G} = \mathcal{L}$ and $\mathcal{C} = \emptyset$ in Theorem 3.2.

Theorem 3.1 is proved in §5. As indicated by part (A), the bulk of the work in that theorem goes into establishing the relationship between X and the countable-state Markov shift Σ, so that strong positive recurrence and liftability can be deduced, relating X to a tower with exponential tails. Once this is done the result follows from pre-existing machinery developed by various people, including especially Sarig [Sar99, Sar01, CS09, Sar11] and Young [You98, You99].

Theorem 3.2 is proved in §6 and uses ideas from the author's previous work with D.J. Thompson on non-uniform specification [CT12, CT13], as well as some new ideas. It is worth pointing out in particular that the construction of $\mathcal{F}$ relies on the following definition, which is illustrated in Figure 3.2 and is inspired by Bertrand's construction of a synchronising word for a shift with specification [Ber88].

Definition 3.4. Given $\mathcal{G} \subset \mathcal{L}$ satisfying **[I]**, we say that (r, c, s) is a **synchronising triple** for $\mathcal{G}$ if $r, s \in \mathcal{G}$, $c \in \mathcal{L}_{\leq \tau}$, and given any $r' \in \mathcal{L}r \cap \mathcal{G}$ and $s' \in s\mathcal{L} \cap \mathcal{G}$, we have $r'cs' \in \mathcal{G}$. In this case we write $\mathcal{B}^{r,s} = \mathcal{L}r \cap s\mathcal{L} \cap \mathcal{G}$, and $\mathcal{F}^{r,c,s} = c\mathcal{B}^{r,s} = \{cw \mid w \in \mathcal{B}^{r,s}\}$.

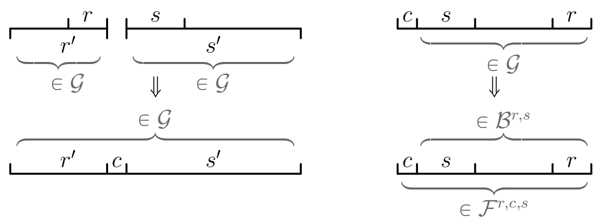


FIGURE 3.2. A synchronising triple (r, c, s) and the collections $\mathcal{B}^{r,s}$, $\mathcal{F}^{r,c,s}$ it generates.

We think of c as the **'connecting'** word and $\mathcal{B}^{r,s}$ as the set of **'bridges'** between instances of c. Note that each $w \in \mathcal{B}^{r,s}$ begins with s and ends with r, but the subwords s and r may overlap; we can have $|w| < |s| + |r|$. We mention some important facts about synchronising triples and the relationship between $\mathcal{G}$ and $\mathcal{F}$; further details are in §6.

Remark 3.5. Every word in $\mathcal{F} = \mathcal{F}^{r,c,s}$ can be extended to a word in $\mathcal{G}$ in a uniform number of symbols, and vice versa. On the one hand, given any $w \in \mathcal{F}$, we have $w = cv$ for some $v \in s\mathcal{L} \cap \mathcal{G}$, hence $rw = rcv \in \mathcal{G}$. On the other hand, if $w \in \mathcal{G}$ then there are $u, v \in \mathcal{L}_{\leq\tau}$ such that $suwvr \in \mathcal{B}^{r,s} = s\mathcal{L} \cap \mathcal{L}r \cap \mathcal{G}$, and hence $csuwvr \in \mathcal{F}$.

Remark 3.6. If $\mathcal{G}$ satisfies $[\mathbf{I}']$, then we can relate $F := \{|w| \mid w \in \mathcal{F}^{r,c,s}\} \subset \mathbb{N}$ and $G := \{|w| \mid w \in \mathcal{G}\} \subset \mathbb{N}$ as in the final claim of Theorem 3.2. Indeed, $|c| = \tau$ implies that every $cw \in \mathcal{F}^{r,c,s}$ has $|cw| = \tau + |w| \in \tau + G$, so $F \subset \tau + G$, and $\gcd(F)$ is a multiple of $d := \gcd(\tau + G)$. On the other hand, $\tau + G \subset \mathbb{N}$ is closed under addition by $[\mathbf{I}']$, so there is $N \in \mathbb{N}$ such that $d\mathbb{N} \cap [N, \infty) \subset \tau + G$. For every $w \in \mathcal{G}$ we have $csuwvr \in \mathcal{F}^{r,c,s}$ for some $u, v \in \mathcal{L}_\tau$, and so $F \supset \tau + G + (2\tau + |r| + |s|) \supset d\mathbb{N} \cap [N + 2\tau + |r| + |s|, \infty)$, which gives $\gcd(F) = d$.

Proposition 3.7. *Every $\mathcal{G} \subset \mathcal{L}$ satisfying* $[\mathbf{I}]$ *has a synchronising triple. Moreover, if (r, c, s) is any synchronising triple for $\mathcal{G}$, then $\mathcal{F}^{r,c,s}$ satisfies* $[\mathbf{I}_0]$. *Finally, a measure μ is Gibbs for φ with respect to $\mathcal{G}$ if and only if it is Gibbs with respect to $\mathcal{F}^{r,c,s}$.*

In §6, we follow an idea of Bertrand [Ber88] to prove Proposition 6.2, which implies Proposition 3.7 and also uses $[\mathbf{III}_\mathrm{a}]$ to find a synchronising triple for which $\mathcal{F}^{r,c,s}$ satisfies $[\mathbf{III}^*]$. The hardest part of Theorem 3.2 is the proof that if $\mathcal{G}$ satisfies $[\mathbf{II}]$, $[\mathbf{III}_\mathrm{a}]$, $[\mathbf{III}_\mathrm{b}]$, then $\mathcal{F}$ satisfies $[\mathbf{II}']$ (Proposition 6.3); this occupies most of §6.

3.2. **Previous results, motivation, and context.**

3.2.1. *Uniform specification.* For shift spaces with the classical specification property, conclusions (i)–(iii) of Theorem 1.1 are well-known [Bow74]. Since shifts with specification are synchronised [Ber88] and have a unique MME, which is fully supported, it follows from results of Thomsen [Tho06] that the

Fischer graph of such a shift is in fact strongly positive recurrent; this could be used to establish conclusions (iv)–(vi) for the unique MME (although this is not done in [Tho06]). It seems likely that this approach could be extended to other equilibrium states, allowing the Fischer cover to replace the construction given in §3.1 here, but this does not appear to have been done in the literature.

Our proof of conclusions (iv)–(vi) uses a Perron–Frobenius theorem and a spectral gap for the appropriate transfer operator on the graph representation of the shift space. Ruelle [Rue92] proved a Perron–Frobenius theorem using specification directly, but did not establish spectral gap or a rate of convergence. One class of shifts with specification for which the transfer operator (for $\varphi = 0$) has been studied explicitly are the *cocyclic subshifts* introduced by Kwapisz [Kwa00, Kwa04]; a cocyclic subshift over a finite alphabet $A = \{1, \dots, m\}$ is defined by fixing a finite dimensional vector space V and linear transformations $\Phi_i \in \mathrm{End}(V)$ for $1 \leq i \leq m$, then declaring a word $w \in A^*$ to be legal if and only if $\Phi_{w_1} \cdots \Phi_{w_{|w|}} \neq 0$. Transitive cocyclic subshifts often fail to be SFTs (or even sofic), but have specification and hence satisfy conclusions (i)–(iii) of Theorem 1.1 by [Bow74]. For the zero potential, spectral properties of the transfer operator were studied in [Kwa04], although conclusions (iv)–(vi) were not discussed there. Theorem 1.1 establishes these conclusions for any Hölder continuous potential on a transitive cocyclic subshift.

3.2.2. *Non-uniform specification.* Condition **[I′]** first appeared in [CT12, CT13], and **[II]** is a version of the following property from [CT13], which appeared in [CT12] in the case $\varphi = 0$.

[II*] There are $\mathcal{C}^p, \mathcal{C}^s \subset \mathcal{L}$ such that $\mathcal{L} \subset \mathcal{C}^p\mathcal{G}\mathcal{C}^s$, and $P(\mathcal{C}^p \cup \mathcal{C}^s, \varphi) < P(\varphi)$.

In [CT12] these two conditions were used to prove existence of a unique measure of maximal entropy, subject to a condition on the collections[12]

$$\begin{aligned} \mathcal{G}^M &= \mathcal{L} \cap (\mathcal{C}^p_{\leq M}\mathcal{G}\mathcal{C}^s_{\leq M}) \\ &= \{uvw \in \mathcal{L} \mid u \in \mathcal{C}^p, v \in \mathcal{G}, w \in \mathcal{C}^s, |u| \leq M, |w| \leq M\}, \end{aligned} \tag{3.5}$$

given by the following **'extendability'** requirement.

[E] For all $M \geq 0$ there is $T = T(M) \in \mathbb{N}$ such that each $w \in \mathcal{G}^M$ has $(\mathcal{L}_{\leq T})w(\mathcal{L}_{\leq T}) \cap \mathcal{G} \neq \emptyset$.

That is, every word in $\mathcal{G}^M$ can be extended to a word in $\mathcal{G}$ by appending at most T symbols to either end. If $\mathcal{G}$ satisfies **[I]** and the sets $\mathcal{G}^M$ from (3.5) satisfy **[E]**, then the following is true.

[I^M] Every $\mathcal{G}^M$ satisfies **[I]** (the gluing time τ may depend on M).

It was shown in [CT13, Theorem C] that if φ is Hölder, then **[I^M]** and **[II*]** imply existence of a unique equilibrium state for φ.

[12]Again we point out that these are not the collections $\mathcal{G}(M)$ in Theorem 1.3.

Remark 3.8. The uniqueness result in [CT13] is stated in the case when $\mathcal{G}^M$ satisfies [$\mathbf{I}'$], but this can be replaced by [$\mathbf{I}$] in both [CT12, CT13]; the proof requires only minor modifications along the lines given in [CT15, Proposition 4.3] (the setting there is more general). A more subtle point is the replacement of [$\mathbf{I}'$] by [$\mathbf{I}$] in the factor results from [CT12]; this is discussed in §3.3.4.

Formally there is little difference between [**II**] and [$\mathbf{II}^*$]: if $\mathcal{C}^p, \mathcal{C}^s$ satisfy [**II**], then one can put $\hat{\mathcal{C}}^p = \mathcal{C}^p \cup (\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s)$ and obtain [$\mathbf{II}^*$] for $\hat{\mathcal{C}}^p, \mathcal{G}, \mathcal{C}^s$. However, this enlarges the collections $\mathcal{G}^M$ and in particular may cause [**E**] and [$\mathbf{I}^M$] to fail. Despite this, an examination of the proof of [CT13, Theorem C] reveals that the result still holds if [$\mathbf{II}^*$] is replaced by [**II**]. Indeed, the only place where [$\mathbf{II}^*$] is used in the proof is in the partition sum estimates in [CT13, §5.1]; Lemma 4.5 of the present paper establishes some of these using the weaker condition [**II**], and the others extend in a completely analogous manner. Together with Remark 3.8, this gives the following.

Theorem 3.9. *Let X be a shift space on a finite alphabet and $\varphi\colon X \to \mathbb{R}$ a Hölder continuous potential.*[13] *Suppose $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}(X)$ are such that*

(1) $\mathcal{G}^M$ satisfies [$\mathbf{I}$] for every $M \in \mathbb{N}$;
(2) $\mathcal{C} := \mathcal{C}^p \cup \mathcal{C}^s \cup (\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s)$ satisfies $P(\mathcal{C}, \varphi) < P(\varphi)$.

Then (X, φ) has a unique equilibrium state μ, which satisfies the Gibbs property (2.11) *with respect to every $\mathcal{G}^M$.*[14] *If $\mathcal{G}$ satisfies [$\mathbf{I}_0$], then μ is the limiting distribution of φ-weighted periodic orbits.*[15]

3.2.3. *Motivating questions; β-shifts, specification, and towers.* The motivating examples for the non-uniform specification property introduced in [CT12] were β-shifts and their factors. Given $\beta > 1$ and $A = \{0, \dots, \lceil \beta \rceil - 1\}$, the greedy β-expansion of 1 is the lexicographically maximal $\mathbf{z} \in A^{\mathbb{N}}$ satisfying $1 = \sum_{k=1}^\infty \mathbf{z}_k \beta^{-k}$, and the β-shift Σ_β is the subshift of $A^{\mathbb{N}}$ defined by the condition that $x \in \Sigma_\beta$ if and only if $x_{[k,\infty)} \preceq \mathbf{z}_{[k,\infty)}$ for all $k \in \mathbb{N}$, where $\preceq$ is the lexicographic order. The β-shift is the natural coding space for the β-transformation $x \mapsto \beta x \pmod 1$, and can be described in terms of a countable state graph; this was done by Hofbauer [Hof78], who used this structure to prove uniqueness of the MME. The corresponding result for factors of β-shifts remained open [Boy08, Problem 28.1], which led the author and D.J. Thompson to introduce conditions [$\mathbf{I}'$], [$\mathbf{II}^*$], and [**E**] in [CT12] as a mechanism for uniqueness that passes to factors. This raised the following two natural questions.

(1) Are there examples of systems with non-uniform specification that do *not* come from a countable graph? To put it very loosely, does

---

[13] We can replace Hölder continuity with the weaker **Bowen property** on $\mathcal{G}$ [CT13, Definition 2.2].

[14] The constant Q_1 in (2.11) is allowed to depend on M.

[15] This last assertion holds because [$\mathbf{I}_0$] implies the (Per)-specification condition from [CT13, Definition 2.1].

the machinery of non-uniform specification apply in a broader setting than the machinery of Young towers?

(2) The tower resulting from Hofbauer's graph structure for the β-shifts (see also [Wal78] for equilibrium states for $\varphi \neq 0$) can be used to get the stronger conclusions (iv)–(vi) that do not follow from [CT12]. Can these conclusions be obtained using non-uniform specification?

These questions motivated the present paper, which says that the answers are "no" and "yes", respectively: replacing **[E]** with **[III]**, non-uniform specification in the sense of **[I]**–**[III]** implies the existence of a Young tower with exponential tails, and thus its statistical consequences are just as strong.

This can be interpreted as a negative result in the sense that we should not expect Theorem 1.1 to cover new classes of examples, since every shift satisfying **[I]**–**[III]** could also be described in terms of a countable graph. This is discussed in §3.3.3.

On the other hand, if the system is defined in a manner that does not make this Markov structure explicit, then it may be difficult to find the graph that does the job, or to determine its properties (consider QFTs or cocyclic subshifts). This is particularly true in the non-symbolic setting, where the task of building a suitable tower can be quite difficult. Thus we can also interpret the above answers as a positive result, since the conditions of non-uniform specification may be easier to verify. In particular, one may hope that a non-symbolic version of Theorem 1.1 will eventually be useful in studying smooth systems, and we discuss this setting next.

3.2.4. *Smooth systems and condition* **[III]**. In [CT14, CT15], the uniqueness results from [CT12, CT13] are generalised to the setting where X is a compact metric space, $f\colon X \to X$ is a continuous map, and $\varphi\colon X \to \mathbb{R}$ is a continuous potential function. An application of these results to non-uniformly hyperbolic diffeomorphisms with dominated splittings is given in [CFT15]. For the smooth systems considered there, it is likely possible to build a Young tower with exponential tails, and thus prove exponential decay of correlations, central limit theorem, etc., for the unique equilibrium state; however, this has not yet been carried out, and work on other similar examples suggests that it presents non-trivial technical challenges, even though the basic ideas are clear [AP10, ADLP16, AL15]. One motivation for the present work is the goal of ultimately giving a set of conditions that imply the existence of such a tower without the need to establish by hand the necessary liftability results and tail decay rate.

In the non-symbolic setting the language $\mathcal{L}$ is replaced with the **space of finite orbit segments** $X \times \mathbb{N}$, where the pair (x, n) is associated to the orbit segment $x, f(x), \dots, f^{n-1}x$. Then one asks the collection $\mathcal{G} \subset X \times \mathbb{N}$ of 'good' orbit segments to satisfy a specification property (among other things). The analogue of condition **[III]** in this setting is as follows.

[III′] There is $L \in \mathbb{N}$ such that if $x \in X$ and $i \leq j \leq k \leq \ell$ are such that $k - j \geq L$ and $(f^j x, \ell - j) \in \mathcal{G}$, then $(f^i x, \ell - i), (f^j x, k - j) \in \mathcal{G}$.

(Note that when (X, f) is a shift space (X, σ), this is equivalent to **[III]**.)

When working with the diffeomorphisms $f: M \to M$ from [CFT15], the idea behind obtaining a collection $\mathcal{G} \subset M \times \mathbb{N}$ with specification is to take the dominated splitting $TM = E^s \oplus E^u$, fix $\chi > 0$, and to let $\mathcal{G}$ be the set of all (x, n) such that

$$\|Df^k|_{E^s(x)}\| \le e^{-\chi k} \text{ for all } 1 \le k \le n, \tag{3.6}$$

$$\|Df^{-k}|_{E^u(f^n x)}\| \le e^{-\chi k} \text{ for all } 1 \le k \le n. \tag{3.7}$$

If $x \in M$ and $i \le j < k \le \ell$ are such that $(f^i x, k-i), (f^j x, \ell - j) \in \mathcal{G}$, then

$$\|Df^a|_{E^s(f^i x)}\| \le e^{-\chi a} \text{ for all } 1 \le a \le k - i, \tag{3.8}$$

$$\|Df^a|_{E^s(f^j x)}\| \le e^{-\chi a} \text{ for all } 1 \le a \le \ell - j. \tag{3.9}$$

It immediately follows that

$$\|Df^a|_{E^s(f^j x)}\| \le e^{-\chi a} \text{ for all } 1 \le a \le k - j;$$

the corresponding observation for E^u shows that $(f^j x, k - j) \in \mathcal{G}$, and a similar proof gives $(f^i x, \ell - i) \in \mathcal{G}$. Thus **[III′]** is automatically satisfied when $\mathcal{G}$ is defined via (3.6)–(3.7).

The basic mechanism at work here continues to make sense in the symbolic setting, where it motivates the conditions of Theorem 1.3. Consider

$$\mathcal{C}^- = \{(x, n) \mid \|Df^n|_{E^s(x)}\| > e^{-\chi n}\}, \tag{3.10}$$

$$\mathcal{C}^+ = \{(x, n) \mid \|Df^{-n}|_{E^u(f^n x)}\| > e^{-\chi n}\}; \tag{3.11}$$

then the collection $\mathcal{G}$ described by (3.6)–(3.7) can be rewritten as

$$\mathcal{G} = \{(x, n) \mid (x, a) \notin \mathcal{C}^- \text{ and } (f^a x, n - a) \notin \mathcal{C}^+ \text{ for all } 1 \le a \le n\},$$

which is a particular case of (1.1). Moreover, we observe that given any $(x, n) \in \mathcal{C}^-$ and $1 \le k \le n$, we have either $(x, k) \in \mathcal{C}^-$ or $(f^k x, n - k) \in \mathcal{C}^-$, which is reminiscent of (1.2) (although a bit weaker).

3.3. **Applications and examples.**

3.3.1. *Shifts of quasi-finite type.* We prove Theorem 1.4 using Theorem 1.3. First we prove (1.2) for $\mathcal{C}^+ = \mathcal{C}^\ell$ and $\mathcal{C}^- = \mathcal{C}^r$. Given $vw \in \mathcal{C}^+ = \mathcal{C}^\ell$, let $u \in \mathcal{L}$ be such that $(vw)_{[2,|vw|]}u \in \mathcal{L}$ but $(vw)u \notin \mathcal{L}$. Then $v_{[2,|v|]}(wu) \in \mathcal{L}$ but $v(wu) \notin \mathcal{L}$, so $v \in \mathcal{C}^+ = \mathcal{C}^\ell$. The proof for $\mathcal{C}^- = \mathcal{C}^r$ is similar.

Now we show that $\mathcal{C}^- = \mathcal{C}^r$ and $\mathcal{C}^+ = \mathcal{C}^\ell$ always form a complete list of obstructions to specification as long as X is topologically transitive. Fix $M \in \mathbb{N}$ and let $\tau \in \mathbb{N}$ be such that for every $v, w \in \mathcal{L}_{\le M}$ there is $u \in \mathcal{L}$ with $|u| \le \tau$ such that $vuw \in \mathcal{L}$; note that such a τ exists because X is transitive and $\mathcal{L}_{\le M}$ is finite. Then given any $v, w \in \mathcal{G}^M(\mathcal{C}^\pm)$, there is $u \in \mathcal{L}_{\le \tau}$ such that $v_{[|v|-M+1,|v|]}uw_{[1,M]} \in \mathcal{L}$. Since $v, w \in \mathcal{G}^M(\mathcal{C}^\pm)$, we have that $v_{[|v|-M,|v|]} \notin \mathcal{C}^+ = \mathcal{C}^\ell$, and hence $v_{[|v|-M,|v|]}uw_{[1,M]} \in \mathcal{L}$. Proceeding inductively and using the fact that $v_{[|v|-i,|v|]} \notin \mathcal{C}^\ell$ for any $i > M$, we conclude

that $vuw_{[1,M]} \in \mathcal{L}$. A similar induction using $w_{[1,i]} \notin \mathcal{C}^- = \mathcal{C}^r$ yields $vuw \in \mathcal{L}$. Since M was arbitrary, this gives [**I***].

The proof for $\mathcal{C}^- = \emptyset$ and $\mathcal{C}^+ = \mathcal{C}^\ell$ in the topologically exact case is similar. Given $M \in \mathbb{N}$, exactness gives $\tau \in \mathbb{N}$ such that for every $v \in \mathcal{L}_{\leq M}$, we have $\sigma^{\tau+M}[v] = X^+$. In particular, for all $v \in \mathcal{L}_{\leq M}$ and $w \in \mathcal{L}$, we have $\sigma^{\tau+M}[v] \supset [w]$, so there is $u \in \mathcal{L}_\tau$ such that $vuw \in \mathcal{L}$. Then given any $v, w \in \mathcal{G}^M(\mathcal{C}^\pm)$, there is $u \in \mathcal{L}_\tau$ such that $v_{(|v|-M,|v|]}uw \in \mathcal{L}$, and the same inductive argument as before shows that $vuw \in \mathcal{L}$, so [**I***] holds.

3.3.2. *Synchronised shifts.* Uniqueness of the MME for synchronised shifts was studied by Thomsen in [Tho06]; these shifts have a canonical presentation via a countable graph (the **Fischer cover**), and Thomsen proved that the corresponding countable-state Markov shift is strongly positive recurrent (for the zero potential) if $h(\partial X) < h(X)$, where ∂X is the **derived shift** consisting of all $x \in X$ that can be approximated by periodic points of X and do not contain any synchronising words. Theorem 1.5 can be viewed as a version of Thomsen's result for non-zero potentials (although the countable graph from §3.1 need not be the Fischer cover).

To deduce Theorem 1.5 from Theorem 1.2, let s be a synchronising word and let $\mathcal{G} = \mathcal{L} \cap s\mathcal{L} \cap \mathcal{L}s$ be the set of words that both start and end with s (note that s is allowed to overlap itself). Choose $c \in \mathcal{L}$ such that $scs \in \mathcal{L}$, and let $\tau = |c|$. Then for every $v, w \in \mathcal{G}$ we have $vcw \in \mathcal{G}$ by the definition of a synchronising word, so $\mathcal{G}$ satisfies [**I**$'$]. Writing $\mathcal{C}^p = \mathcal{C}^s = \mathcal{L}(Y) = \mathcal{L} \setminus \mathcal{L}s\mathcal{L}$ for the collection of words that do not contain s as a subword, we see that every $w \in \mathcal{L}$ is either contained in $\mathcal{C}^p$, or can be decomposed as $u^p v u^s \in \mathcal{C}^p\mathcal{G}\mathcal{C}^s$ by marking the first and last occurrences of s as a subword of w. Thus $P(\mathcal{C}^p \cup \mathcal{C}^s \cup (\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s), \varphi) = P(Y, \varphi)$, and the hypothesis that $P(Y, \varphi) < P(\varphi)$ implies [**II**]. Finally, taking $L = |s|$ we see that [**III**] is immediately satisfied by the definition of $\mathcal{G}$.

3.3.3. *Coded shifts.* Every system with the uniform specification property is synchronised, and hence coded, so it is natural to ask whether systems with the non-uniform specification property given by [**I**]–[**III**] lie in these classes.

One quickly sees that such systems need not be synchronised; indeed, every β-shift satisfies [**I**]–[**III**], but not all β-shifts are synchronised [Sch97].

On the other hand, Theorem 1.2 shows that [**I**]–[**III**] imply that X contains the coded shift $X' := \overline{\pi(X)}$ (with uniquely decipherable generating set $\mathcal{F}$), and that every equilibrium state μ for (X, σ, φ) has $\mu(X') = 1$; thus from the thermodynamic point of view, every question about a system satisfying [**I**]–[**III**] is a question about a coded shift. We point out, though, that we may not have $X' = X$ even if X has [**I**]–[**III**]. For example, if $X \subset \{0,1,2\}^\mathbb{Z}$ is the SFT defined by forbidding the words 20 and 21, then X satisfies [**I**]–[**III**] for $\varphi = 0$ by taking $\mathcal{F} = \{0,1\}^*$, $\mathcal{C}^p = \emptyset$, and $\mathcal{C}^s = \{2\}^*$, but the corresponding coded subshift is $X' = \{0,1\}^\mathbb{Z} \neq X$.

Given a coded shift X with language $\mathcal{L}$ and generating set G, we saw already that a natural way to approach the thermodynamic properties is to consider the collection I of irreducible generators and then let $\mathcal{D} = \{w_{[i,j]} \mid w \in I, 1 \le i \le j \le |w|\}$; putting $\mathcal{F} = I^*$ and

$$\mathcal{E}^p = \{w_{[i,|[|w|]]} \mid w \in I, 1 \le i \le |w|\}, \quad \mathcal{E}^s = \{w_{[1,i]} \mid w \in I, 1 \le i \le |w|\},$$

we get $I \cup \mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s) \subset \mathcal{D}$, so that in particular if $P(\mathcal{D}, \varphi) < P(\varphi)$ then we have [**II**′]. If I is uniquely decipherable then we can apply Theorem 3.1 and deduce the conclusions of Theorems 1.1 and 1.2. However, even without unique decipherability we can still apply Theorem 3.9; it is shown in [CT12, §4] that with this choice of $\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s$, condition [**E**] holds, and in particular, every $\mathcal{G}^M$ satisfies [**I**]. Thus by Theorem 3.9, we have proved the following.[16]

Theorem 3.10. *Let X be a coded shift and $\varphi\colon X \to \mathbb{R}$ Hölder continuous. If there is a generating set I for X such that $P(\mathcal{D}(I), \varphi) < P(\varphi)$, then*

(1) (X, φ) has a unique equilibrium state μ;
(2) μ has the Gibbs property (2.11) with respect to $\mathcal{F} = I^*$;
(3) μ is the limiting distribution of φ-weighted periodic orbits.

If in addition we know that I is uniquely decipherable, then we can apply Theorem 3.1 and deduce that μ satisfies the stronger statistical properties (iv)–(vi) as well; these do not follow from Theorem 3.9. Without unique decipherability, we can still obtain X as $\overline{\pi(\Sigma)}$, but since we have no information on the multiplicity of the map π, we cannot deduce that Σ is strongly positive recurrent. Moreover, it is possible that π decreases entropy.

Example 3.11. Let $X \subset \{0,1\}^{\mathbb{Z}}$ be the SFT defined by forbidding the word 111. Let $\mathcal{F} \subset \mathcal{L}(X)$ be the set of all words that neither start nor end with the word 11. Then $\mathcal{F}$ satisfies condition [$\mathbf{I}_0$], and we get $I = I(\mathcal{F}) = \{0, 01, 10\}$ as the irreducible elements of $\mathcal{F}$. But then $010 = (01)(0) = (0)(10) \in \mathcal{F}$ has two different ‘factorisations’, which can also be used to show that [**III***] fails and $\pi^{-1}((010)^\infty)$ is uncountable. It turns out that $h(\Sigma) = \log 2 > h(X)$.

It is shown in [BH86, Proposition 2.1] that every coded shift admits a uniquely decipherable generating set, which can be used to build a good cover Σ; similarly, [FF92, Theorem 1.7] shows that it is always possible to build a **‘bi-resolving’** cover, which in particular gives a 1-1 map π. However, in both cases one must abandon the original generating set I and pass to a new generating set I', for which the set of obstructions $\mathcal{D}(I')$ may be quite large, and in particular there is no a priori reason why [**II**′] should hold. Since [**II**′] was required to prove strong positive recurrence, it is not clear whether one can get a proof of statistical properties in this way. Thus we have the following open question.

[16] This result corrects an error in [CT12, Theorem B], where the case $\varphi = 0$ was considered and $h(\mathcal{E}^p \cup \mathcal{E}^s)$ was used instead of $h(\mathcal{D})$. The problem there was that we may have $\mathcal{L} \not\subset \mathcal{E}^p \mathcal{F} \mathcal{E}^s$; this motivates the more general version of [**II**] used in this paper.

Question 3.12. Let X be a coded shift with a (not uniquely decipherable) generating set $I \subset \mathcal{L}(X)$ such that $P(\mathcal{D}(I), \varphi) < P(\varphi)$ for some Hölder φ. Let μ be the unique equilibrium state for (X, φ) guaranteed by Theorem 3.10. Do conclusions (iv)–(vi) of Theorem 1.1 still hold? That is, is some iterate of (X, σ, μ) Bernoulli with exponential decay of correlations; does (X, σ, μ) satisfy the central limit theorem; and is the pressure function real analytic at φ?

As observed above, our results reduce thermodynamic questions for shifts with **[I]**–**[III]** to questions about a coded shift with uniquely decipherable generating set I for which $\mathcal{F} = I^*$ satisfies **[II′]**. However, it is not clear whether the collection $\mathcal{D} = \mathcal{D}(I)$ has $P(\mathcal{D}, \varphi) < P(\varphi)$.

Question 3.13. Let X be a coded shift with uniquely decipherable generating set I satisfying **[II′]**: there are $\mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L}(X)$ such that $P(I \cup \mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p I^* \mathcal{E}^s), \varphi) < P(\varphi)$. Must X have a uniquely decipherable generating set I' for which $P(\mathcal{D}(I'), \varphi) < P(\varphi)$?

3.3.4. *Factors.* Before stating our results for factors we describe how Theorem 1.3 applies to the motivating examples from [CT12, CT13]: β-shifts (defined in §3.2.3) and S-gap shifts.

Example 3.14. For the β-shift, let $\mathbf{z}$ be the β-expansion of 1 and take $\mathcal{C}^+ = \{\mathbf{z}_{[0,n)} \mid n = 0, 1, 2, \dots\}$ and $\mathcal{C}^- = \emptyset$, so $\mathcal{C}^\pm$ satisfy (1.2). In the standard graph presentation of the β-shift, $\mathcal{G}(\mathcal{C}^\pm, M)$ is the collection of words that label paths starting at the base vertex and ending in the first M vertices, and so it satisfies **[I*]** for the reasons discussed in [CT12]. Thus $\mathcal{C}^\pm$ is a complete list of obstructions to specification in the sense of Theorem 1.3.

Example 3.15. Given an infinite subset $S \subset \mathbb{N} \cup \{0\}$, the S-gap shift X_S is the coded system with generating set $\{10^n \mid n \in S\}$. Take $\mathcal{C}^+ = \mathcal{C}^- = \{0^k \mid k \in \mathbb{N}\}$, then $\mathcal{C}^\pm$ satisfy (1.2), and we have $\mathcal{G}(\mathcal{C}^\pm, M) = \{0^a 1 w 1 0^b \in \mathcal{L} \mid a, b \leq M\}$. Then taking $\tau(M) = 2 \min\{s \in S \mid s \geq M\}$, we see that any two words in $\mathcal{G}(\mathcal{C}^\pm, M)$ can be joined by 0^c for some $0 \leq c \leq \tau(M)$, and so **[I*]** holds. Thus $\mathcal{C}^\pm$ is a complete list of obstructions to specification.

It is shown in [CT13] and [CTY, §5.1.3] that for both examples above $P(\mathcal{C}^\pm, \varphi) < P(\varphi)$ for every Hölder φ; here we restrict our attention to $\varphi = 0$ and define the **entropy of obstructions to specification** to be[17]

$$h^\perp_{\mathrm{spec}}(X) = \inf\{h(\mathcal{C}^\pm) \mid \mathcal{C}^\pm \subset \mathcal{L}(X) \text{ satisfy (1.2) and } [\mathbf{I}^*]\}. \tag{3.12}$$

Note that $h^\perp_{\mathrm{spec}}(X) = 0$ for both β-shifts and S-gap shifts. Although the most obvious way to get $h^\perp_{\mathrm{spec}}(X) = 0$ is to have $h(\mathcal{C}^\pm) = 0$ for some $\mathcal{C}^\pm$, we expect that there are examples where $h^\perp_{\mathrm{spec}}(X) = 0$ but the infimum is not

[17]A similar quantity was defined in [CT14]; this differs in that we consider (1.2) and **[I*]** instead of **[I^M]**.

achieved; a natural class of candidates is given by shift spaces coding transitive piecewise monotonic transformations of the interval, whose structure has been described by Hofbauer [Hof79, Hof81].

It follows from Theorem 1.3 that $h^\perp_{\mathrm{spec}}(X) < h(X)$ implies existence of a unique MME together with the other conclusions of Theorem 1.1. Moreover, $h^\perp_{\mathrm{spec}}$ is non-increasing under factors; the following is proved in §7.

Proposition 3.16. *If $\tilde X$ is a shift factor of X, then $h^\perp_{\mathrm{spec}}(\tilde X) \leq h^\perp_{\mathrm{spec}}(X)$.*

In particular, if $h^\perp_{\mathrm{spec}}(X) = 0$, then every shift factor of X also has $h^\perp_{\mathrm{spec}} = 0$. We also consider shift spaces with the following property.

[$\mathbf{E}^*$] There is $\mathcal{G} \subset \mathcal{L}$ with [**I**] s.t. every $w \in \mathcal{L}$ has $u, v \in \mathcal{L}$ with $uwv \in \mathcal{G}$.

Note that we do not require $\mathcal{G}$ in [$\mathbf{E}^*$] to satisfy [**III**]; in particular, it does not need to be the same collection produced by Theorem 1.3. The following is proved in §7.

Proposition 3.17. *If X satisfies [$\mathbf{E}^*$], then every subshift factor $\tilde X$ of X has $h(\tilde X) > 0$ or is a single periodic orbit. In particular, if X satisfies [$\mathbf{E}^*$] and $\gcd\{k \mid \mathrm{Per}_k(X) \neq \emptyset\} = 1$, then every non-trivial subshift factor of X has positive entropy.*

Propositions 3.16 and 3.17 combine with Theorem 1.3 to prove the following result, which is similar to [CT12, Corollary 2.3 and Theorem D] but has stronger conclusions.

Theorem 3.18. *Let (X, σ) be a shift space on a finite alphabet.*

(1) Let $(\tilde X, \tilde\sigma)$ be a subshift factor of (X, σ) such that $h(\tilde X) > h^\perp_{\mathrm{spec}}(X)$. Then $\tilde X$ has a unique measure of maximal entropy, which is the limiting distribution of periodic orbits, has the Bernoulli property and exponential decay of correlations up to a period, and satisfies the central limit theorem; moreover, for any Hölder $\psi \colon \tilde X \to \mathbb{R}$ there is $\varepsilon > 0$ such that $t \mapsto P(t\psi)$ is real analytic on $(-\varepsilon, \varepsilon)$.

(2) Suppose X satisfies [$\mathbf{E}^*$], $\gcd\{k \mid \mathrm{Per}_k(X) \neq \emptyset\} = 1$, and $h^\perp_{\mathrm{spec}}(X) = 0$. Then every subshift factor of (X, σ) satisfies the conclusion of the previous part.

Corollary 3.19. *Let $\tilde X$ be a nontrivial subshift factor of a β-shift or of an S-gap shift. Then $\tilde X$ has a unique MME μ; moreover, μ is the limiting distribution of periodic orbits, has the Bernoulli property and exponential decay of correlations up to a period, and satisfies the central limit theorem. Finally, $t \mapsto P(t\psi)$ is real analytic on a neighbourhood of 0 for every Hölder $\psi \colon \tilde X \to \mathbb{R}$.*

Proof. If S is finite then $\tilde X$ is sofic and has specification, so let X be a β-shift or an S-gap shift with S infinite. Examples 3.14 and 3.15 give $h^\perp_{\mathrm{spec}}(X) = 0$. Since X contains the sequences of all 0s, which is a fixed point, we have $\gcd\{k \mid \mathrm{Per}_k(X) \neq \emptyset\} = 1$. Finally, the collection $\mathcal{G} \subset \mathcal{L}$ from Examples 3.14 and 3.15 satisfies [$\mathbf{E}^*$], and the conclusion follows from Theorem 3.18. ∎

3.3.5. *Hyperbolic potentials.* Since Theorem 3.18 only deals with measures of maximal entropy, one may ask what can be said about equilibrium states for non-zero potentials on the factors $(\tilde{X}, \tilde{\sigma})$. Following [IRRL12], say that a potential $\varphi\colon X \to \mathbb{R}$ is **hyperbolic** if

$$\overline{\lim_{n\to\infty}} \sup_{x\in X} \frac{1}{n} S_n\varphi(x) < P(\varphi). \tag{3.13}$$

If φ is hyperbolic and $h^{\perp}_{\text{spec}}(X) = 0$, then a simple computation shows that there are $\mathcal{C}^{\pm}$ satisfying (1.2), [$\mathbf{I}^*$], and $P(\mathcal{C}^- \cup \mathcal{C}^+, \varphi) < P(\varphi)$; in particular, we can apply Theorem 1.3 as long as φ is Hölder.

SFTs have the property that every Hölder potential is hyperbolic. Buzzi proved that the same is true for the coding spaces of continuous topologically transitive piecewise monotonic interval maps [Buz04], and conjectured that the result remains true without the assumption of continuity. The result is known for a broad class of non-uniformly expanding interval maps [LRL14], and for β-shifts [CT13, Proposition 3.1] and S-gap shifts [CTY, (5.1)]. The proofs of this result for β-shifts and for S-gap shifts are very specific to these examples and in particular do not pass to their factors.

On the other hand, there are coded systems for which some Hölder potentials are not hyperbolic: for example, if X is the shift generated by $\{0^n 1^n \mid n \in \mathbb{N}\}$ and $\varphi = t\mathbf{1}_{[0]}$ for $|t|$ sufficiently large, then (3.13) fails and φ is not hyperbolic [Con].

Question 3.20. Is there an axiomatic condition on a shift space X, weaker than specification (perhaps some form of non-uniform specification), guaranteeing that every Hölder potential on X is hyperbolic? Is there such a condition that is preserved under passing to factors? In particular, does every subshift factor of a β-shift or an S-gap shift have the property that every Hölder potential is hyperbolic?

4. Preparation for the proofs

Before proving Theorems 3.1 and 3.2 in §§5–6, we establish some preparatory results that will be needed later on. In §4.1 we prove stronger versions of the Birkhoff and Shannon–McMillan–Breiman ergodic theorems (Theorems 4.1 and 4.2). In §4.2 we show that Hölder potentials have bounded distortion within cylinders, and use this together with [**I**] and [**II**] to establish a number of uniform bounds on partition sums, a process that also played a key role in [CT12, CT13].

4.1. **Ergodic theorems.** In the proof of Theorem 3.1, we will need mild strengthenings of the Birkhoff and Shannon–McMillan–Breiman ergodic theorems. These are general results that hold beyond the setting of this paper.

Recall that the usual version of the Birkhoff ergodic theorem [Pet89, Theorem 2.2.3] can be stated as follows: if (X, T, μ) is an ergodic measure-preserving transformation and $f\colon X \to \mathbb{R}$ is an L^1 function, then for μ-a.e.

$x \in X$ and every $\varepsilon > 0$ there is $N = N(x, \varepsilon)$ such that for all $n \geq N$ we have $\left|\frac{1}{n} S_n f(x) - \int f\, d\mu\right| < \varepsilon$. We will prove the following stronger version.

Theorem 4.1. *If (X, T, μ) is an invertible ergodic measure-preserving transformation and $f\colon X \to \mathbb{R}$ is an L^1 function, then for μ-a.e. $x \in X$ and every $\varepsilon > 0$ there is $N = N(x, \varepsilon)$ such that for all $n \geq N$ and $\ell \in [0, n]$ we have*

$$\left|\frac{1}{n} S_n f(T^{-\ell} x) - \int f\, d\mu\right| < \varepsilon. \tag{4.1}$$

Given (X, T, μ) as above and a countable (or finite) measurable partition α of X, write $\alpha(x)$ for the partition element containing x, and for $i < j \in \mathbb{Z}$, write $\alpha_i^j = \bigvee_{k=i}^{j-1} T^{-k}\alpha$. Recall that

$$H_\mu(\alpha) := \sum_{A \in \alpha} -\mu(A) \log \mu(A), \qquad h_\mu(\alpha, T) := \lim_{n \to \infty} \frac{1}{n} H_\mu(\alpha_0^n).$$

The Shannon–McMillan–Breiman theorem [Pet89, Theorem 6.2.3] states that if α has $H_\mu(\alpha) < \infty$, then for μ-a.e. $x \in X$ and every $\varepsilon > 0$ there is $N = N(x, \varepsilon)$ such that for all $n \geq N$ we have $\left|-\frac{1}{n} \log \mu(\alpha_0^n(x)) - h_\mu(\alpha, T)\right| < \varepsilon$. We will prove the following stronger version.

Theorem 4.2. *If (X, T, μ) is an invertible ergodic measure-preserving transformation and α is a countable measurable partition with $H_\mu(\alpha) < \infty$, then for μ-a.e. $x \in X$ and every $\varepsilon > 0$ there is $N = N(x, \varepsilon)$ such that for all $n \geq N$ and $\ell \in [0, n]$ we have*

$$\left|-\frac{1}{n} \log \mu(\alpha_{-\ell}^{n-\ell}(x)) - h_\mu(\alpha, T)\right| < \varepsilon. \tag{4.2}$$

The following lemmas will be needed in the proofs of Theorems 4.1 and 4.2. Let (X, T, μ) be an invertible ergodic measure-preserving transformation. Given an L^1 function $f\colon X \to \mathbb{R}$, we write $\tilde{f}_n(x) = \max\{|f(T^k x)| \mid -n \leq k \leq n\}$.

Lemma 4.3. *For every $f \in L^1$ and μ-a.e. x, we have $\lim_{n\to\infty} \frac{1}{n} \tilde{f}_n(x) = 0$.*

Proof. By Birkhoff's ergodic theorem, μ-a.e. x is such that $\frac{1}{n} \sum_{k=0}^{n-1} f(T^k x)$ and $\frac{1}{n} \sum_{k=0}^{n-1} f(T^{-k} x)$ both converge; thus each such x has $\frac{1}{n} f(T^n x) \to 0$ and $\frac{1}{n} f(T^{-n} x) \to 0$.

Let $k(n) \in [-n, n]$ be such that $\tilde{f}_n(x) = f(T^{k(n)} x)$. Note that $|k(n)|$ is non-decreasing. If $|k(n)| \to \infty$ then we have

$$\left|\frac{\tilde{f}_n(x)}{n}\right| = \left|\frac{f(T^{k(n)} x)}{k(n)}\right| \frac{|k(n)|}{n} \to 0$$

since $|k(n)| \leq n$ and $\frac{f(T^{k(n)} x)}{k(n)} \to 0$ (using the fact that $|k(n)| \to \infty$). If on the other hand the sequence $|k(n)|$ is bounded, say by k', then we have $\frac{1}{n} \left|\tilde{f}_n(x)\right| \leq \frac{1}{n} \left|\tilde{f}_{k'}(x)\right| \to 0$. $\square$

Lemma 4.4. *Let (X, T, μ) be an ergodic measure-preserving transformation and $f \in L^1(X, \mu)$. Then for μ-a.e. $x \in X$, for every $\varepsilon > 0$ there is $\delta > 0$ such that $\left|\frac{1}{n}\sum_{k=0}^{m-1} f(T^k x)\right| < \varepsilon$ whenever $1 \leq m \leq \delta n$.*

Proof. Let $x \in X$ be such that $f(T^k x)$ is finite for every $k \in \mathbb{Z}$, and such that $\frac{1}{n} S_n f(x) \to \int f\, d\mu$. Then there is $m_0 \in \mathbb{N}$ such that for all $m \geq m_0$ we have $|S_m f(x)| \leq 2m \int |f|\, d\mu$. Let $M = \max\{|S_m f(x)| \mid 1 \leq m < m_0\}$. Thus for every $m \geq 1$ we have $S_m f(x) \leq \max(2m\|f\|_1, M)$. Now for every $\delta > 0$ and $1 \leq m \leq \delta n$, we have

$$\frac{1}{n}\sum_{k=0}^{m-1} f(T^k x) \leq \max\left(2\frac{m}{n}\|f\|_1, \frac{M}{n}\right) \leq \max(2\delta\|f\|_1, M\delta).$$

Choosing $\delta > 0$ such that this quantity is $< \varepsilon$ completes the proof. □

Proof of Theorem 4.1. Fix $\varepsilon > 0$. Applying the Birkhoff ergodic theorem to f for T and T^{-1}, for μ-a.e. $x \in X$ there is $N = N(x, \varepsilon)$ such that for all $n \geq N$ we have

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} f(T^k x) - \int f\, d\mu\right| < \frac{\varepsilon}{3}, \qquad \left|\frac{1}{n}\sum_{k=1}^{n} f(T^{-k} x) - \int f\, d\mu\right| < \frac{\varepsilon}{3}. \tag{4.3}$$

Applying Lemma 4.4 to T and T^{-1}, for μ-a.e. $x \in X$ there is $\delta > 0$ (depending on x) such that $\delta\|f\|_1 < \varepsilon/3$ and such that for every $m, n \in \mathbb{N}$ with $m \leq \delta n$, we have

$$\left|\frac{1}{n}\sum_{k=1}^{m} f(T^k x)\right| < \frac{\varepsilon}{3}, \qquad \left|\frac{1}{n}\sum_{k=1}^{m} f(T^{-k} x)\right| < \frac{\varepsilon}{3}. \tag{4.4}$$

Let $N' = N(x, \varepsilon)/\delta$, and suppose we have $n \geq N'$ and $\ell \in [0, n]$. We want to estimate $\sum_{k=-\ell}^{-\ell+n-1} f(T^k x) = \sum_{k=1}^{\ell} f(T^{-k} x) + \sum_{k=0}^{n-\ell-1} f(T^k x)$. There are three cases to consider: $\ell \in [0, N]$, $\ell \in (N, n - N)$, and $\ell \in [n - N, n]$.

In the second case we can apply (4.3) to get

$$\begin{aligned}\left|\sum_{k=-\ell}^{-\ell+n-1} f(T^k x) - n\int f\, d\mu\right| &\leq \left|\sum_{k=1}^{\ell} f(T^{-k} x) - \ell\int f\, d\mu\right| \\ &\qquad + \left|\sum_{k=0}^{n-\ell-1} f(T^k x) - (n-\ell)\int f\, d\mu\right| \\ &\leq \ell\tfrac{\varepsilon}{3} + (n-\ell)\tfrac{\varepsilon}{3} = n\tfrac{\varepsilon}{3} < n\varepsilon.\end{aligned}$$

In the first and third cases we must use one of the inequalities from (4.3) and one from (4.4); for example, if $0 \leq \ell \leq N \leq \delta N' \leq \delta n$, then we have

$$\begin{aligned}\left|\sum_{k=-\ell}^{-\ell+n-1} f(T^k x) - n\int f\, d\mu\right| &\leq \left|\sum_{k=1}^{\ell} f(T^{-k} x)\right| + \ell\|f\|_1 + (n-\ell)\frac{\varepsilon}{3} \\ &< n\tfrac{\varepsilon}{3} + \delta n\|f\|_1 + n\tfrac{\varepsilon}{3} \leq n\varepsilon.\end{aligned}$$

The case $\ell \in [n - N, n]$ is analogous. This proves Theorem 4.1. □

Proof of Theorem 4.2. Again, let (X, T, μ) be an invertible ergodic measure-preserving transformation, and let α be a countable measurable partition of X with $H_\mu(\alpha) < \infty$. We want to prove that for μ-a.e. $x \in X$ and every $\varepsilon > 0$ there is $N = N(x, \varepsilon)$ such that for all $n \geq N$ and all $\ell \in [0, n]$ we have

$$\left| -\tfrac{1}{n} \log \mu(\alpha_{-\ell}^{-\ell+n}(x)) - h_\mu(\alpha, T) \right| < \varepsilon. \tag{4.5}$$

We show that this follows from Theorem 4.1, using the standard argument for proving the Shannon–McMillan–Breiman theorem from the Birkhoff ergodic theorem. We follow the presentation in [Pet89, Theorem 6.2.3].

Let $f_n(x) = -\log\left(\frac{\mu(\alpha_0^n(x))}{\mu(\alpha_1^n(x))}\right)$, and $f^* = \sup_{n\geq 1} f_n$. Then $f^* \in L^1$ [Pet89, Corollary 6.2.2], and as shown in the proof of [Pet89, Theorem 6.2.3], we have $f_n \to f$ both pointwise a.e. and in L^1, where f is an L^1 function such that $\int f\, d\mu = h(\alpha, T)$.

Let $I_n(x) = -\log \mu(\alpha_0^n(x))$, so that $f_n = I_n - I_{n-1} \circ T$. We see that

$$I_n = f_n + I_{n-1} \circ T = f_n + f_{n-1} \circ T + I_{n-2} \circ T^2 = \cdots = \sum_{k=0}^{n-1} f_{n-k} \circ T^k.$$

Thus

$$\frac{1}{n} I_n(x) = \frac{1}{n} S_n f(x) + \frac{1}{n} \sum_{k=0}^{n-1} (f_{n-k} - f)(T^k x). \tag{4.6}$$

Note that $-\log \mu(\alpha_{-\ell}^{-\ell+n}(x)) = I_n(T^{-\ell}x)$, so (4.5) can be rewritten as

$$\left| \frac{1}{n} I_n(T^{-\ell}x) - h_\mu(\alpha, T) \right| < \varepsilon. \tag{4.7}$$

We can use (4.6) to get

$$\frac{1}{n} I_n(T^{-\ell}x) = \frac{1}{n} S_n f(T^{-\ell}x) + \frac{1}{n} \sum_{k=0}^{n-1} (f_{n-k} - f)(T^{k-\ell}x). \tag{4.8}$$

By Theorem 4.1, for μ-a.e. $x \in X$ and every $\varepsilon > 0$ there is N such that for all $n \geq N$ we have

$$\left| \frac{1}{n} S_n f(T^{-\ell}x) - h_\mu(\alpha, T) \right| = \left| \frac{1}{n} S_n f(T^{-\ell}x) - \int f\, d\mu \right| \leq \frac{\varepsilon}{2}.$$

Thus to prove (4.7) it suffices to show that there is N' such that for all $n \geq N'$ and $0 \leq \ell \leq n$ we have

$$\frac{1}{n} \sum_{k=0}^{n-1} \left| (f_{n-k} - f)(T^{k-\ell}x) \right| < \frac{\varepsilon}{2}. \tag{4.9}$$

Given $m \in \mathbb{N}$, let $F_m = \sup_{k \geq m} |f_k - f|$. Then

$$\begin{aligned} \frac{1}{n}\sum_{k=0}^{n-1}\left|(f_{n-k}-f)(T^{k-\ell}x)\right| &= \frac{1}{n}\sum_{k=0}^{n-m}\left|(f_{n-k}-f)(T^{k-\ell}x)\right| \\ &\quad + \frac{1}{n}\sum_{k=n-m+1}^{n-1}\left|(f_{n-k}-f)(T^{k-\ell}x)\right|. \end{aligned} \tag{4.10}$$

Because $|f_{n-k} - f| \leq f^* + f \in L^1$ for all k, we can control the second term by applying Lemma 4.3 to $g := f^* + f$, observing that

$$\begin{aligned} \frac{1}{n}\sum_{k=n-m+1}^{n-1}\left|(f_{n-k}-f)(T^{k-\ell}x)\right| &\leq \frac{1}{n}\sum_{k=n-m+1}^{n-1}(f^*+f)(T^{k-\ell}x) \\ &\leq \frac{m}{n}\tilde{g}_n(x) \to 0 \end{aligned}$$

as $n \to \infty$ since m is fixed. For the first term in (4.10), we observe that

$$\frac{1}{n}\sum_{k=0}^{n-m}\left|(f_{n-k}-f)(T^{k-\ell}x)\right| \leq \frac{1}{n}\sum_{k=0}^{n-m}F_m(T^{k-\ell}x),$$

and since $0 \leq F_m \leq f^* + f \in L^1$, we can apply the strengthened ergodic theorem from the previous section, showing that for μ-a.e. x, this is bounded above by $2\int F_m\,d\mu$ for all sufficiently large n. By the dominated convergence theorem and the fact that $F_m \to 0$ pointwise a.e., this can be made arbitrarily small by taking m sufficiently large. □

4.2. **Bounded distortion and counting estimates.** Given $\beta > 0$ and $\varphi \in C_\beta(X)$, we see that for every $w \in \mathcal{L}_n$ and every $x, y \in [w]$, (2.4) yields

$$\left|\varphi(\sigma^k x) - \varphi(\sigma^k y)\right| \leq |\varphi|_\beta\, e^{-\beta \min(k, n-k)} \text{ for all } 0 \leq k < n,$$

so that in particular

$$|S_n\varphi(x) - S_n\varphi(y)| \leq 2\,|\varphi|_\beta \sum_{j=0}^{\infty} e^{-\beta j} =: |\varphi|_d < \infty. \tag{4.11}$$

This can be thought of as a bounded distortion condition, and we think of $|\varphi|_d$ as the bound on distortion of $S_n\varphi$ within an n-cylinder.

It follows from (4.11) that for every $v, w \in \mathcal{L}$ such that $vw \in \mathcal{L}$, we have

$$\hat{\varphi}(v) + \hat{\varphi}(w) - |\varphi|_d \leq \hat{\varphi}(vw) \leq \hat{\varphi}(v) + \hat{\varphi}(w). \tag{4.12}$$

(The upper bound is immediate from (2.1).)

In the proofs of both Theorems 3.1 and 3.2, we will need various estimates on partition sums over $\mathcal{L}$ and over $\mathcal{G}$. We start with the general observation that given $\mathcal{C}, \mathcal{D} \subset \mathcal{L}$ and $m, n \in \mathbb{N}$, the bound $\hat{\varphi}(uv) \leq \hat{\varphi}(u) + \hat{\varphi}(v)$ gives

$$\Lambda_{m+n}(\mathcal{C}_m\mathcal{D}_n, \varphi) \leq \sum_{u \in \mathcal{C}_m}\sum_{v \in \mathcal{D}_n} e^{\hat{\varphi}(u)} e^{\hat{\varphi}(v)} = \Lambda_m(\mathcal{C}, \varphi)\Lambda_n(\mathcal{D}, \varphi); \tag{4.13}$$

this will be used in several places.

The following are similar to estimates appearing in [CT12, Lemmas 5.1–5.4] and [CT13, Section 5]. The chief difference here is that we may have $\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s \neq \emptyset$, but because the pressure of this collection is controlled, we get the same results.

Lemma 4.5. *Let X be a shift space on a finite alphabet and $\varphi \in C_\beta(X)$ for some $\beta > 0$. Let $\mathcal{G} \subset \mathcal{L}(X)$ be such that* **[I]** *and* **[II]** *hold. Then there is $Q_2 > 0$ such that for every n we have*

$$e^{nP(\varphi)} \leq \Lambda_n(\varphi) \leq Q_2 e^{nP(\varphi)}. \tag{4.14}$$

Furthermore, there is $Q_3 > 0$ and $N \in \mathbb{N}$ such that for every sufficiently large n there is $j \in (n - N, n]$ with

$$\Lambda_j(\mathcal{G}, \varphi) \geq Q_3 e^{jP(\varphi)}. \tag{4.15}$$

Proof. For the first inequality in (4.14), we observe that $\mathcal{L}_{kn} \subset \mathcal{L}_n\mathcal{L}_n \cdots \mathcal{L}_n$ (k times), and so by iterating (4.13) we get

$$\Lambda_{kn}(\varphi) \leq \Lambda_n(\varphi)^k, \tag{4.16}$$

which yields $\frac{1}{kn}\log\Lambda_{kn}(\varphi) \leq \frac{1}{n}\log\Lambda_n(\varphi)$. Sending $k \to \infty$ gives the first half of (4.14). Next we prove

$$\Lambda_n(\mathcal{G}, \varphi) \leq Ce^{nP(\varphi)} \tag{4.17}$$

using **[I]** and (4.11), then use this together with **[II]** to prove the second half of (4.14).

By **[I]** there is a map $\pi\colon \mathcal{G}_m \times \mathcal{G}_n \to \mathcal{G}$ given by $\pi(v, w) = vuw$, where $u \in \mathcal{L}$ depends on v, w but always satisfies $|u| \leq \tau$. Iterating and abusing notation slightly gives a map $\pi\colon (\mathcal{G}_n)^k \to \mathcal{G}$ of the form $\pi(v^1, \dots, v^k) = v^1u^1v^2 \cdots u^{k-1}v^k$. Truncating this image to the first kn symbols gives a map $\hat{\pi}\colon (\mathcal{G}_n)^k \to \mathcal{L}_{nk}$. Because we delete at most $k\tau$ symbols to go from π to $\hat{\pi}$, we have $\#\hat{\pi}^{-1}(w) \leq (\#A + 1)^{k\tau}$. (We need $\#A + 1$ instead of $\#A$ since the number of deleted symbols is at most $k\tau$, rather than exactly $k\tau$.)

Furthermore, (4.12) yields

$$\hat{\varphi}(v^1u^1 \cdots u^{k-1}v^k) \geq \hat{\varphi}(v^1) + \cdots + \hat{\varphi}(v^k) - k(\tau\|\varphi\| + |\varphi|_d),$$

and since truncation deletes at most $k\tau$ symbols, we have

$$\hat{\varphi}(\hat{\pi}(v^1, \dots, v^k)) \geq \left(\sum_{i=1}^{k} \hat{\varphi}(v^i)\right) - C'k.$$

It follows that $\Lambda_{kn}(\varphi) \geq (\#A + 1)^{-k\tau} e^{-C'k} \Lambda_n(\mathcal{G}, \varphi)^k$, hence

$$\frac{1}{kn}\log\Lambda_{kn}(\varphi) \geq \frac{1}{n}\log\Lambda_n(\mathcal{G}, \varphi) - \frac{1}{n}(C' + \tau\log(\#A + 1)).$$

Sending $k \to \infty$ gives (4.17).

By Condition **[II]** there is $\varepsilon > 0$ and $K > 0$ such that

$$\Lambda_n(\mathcal{C}^p \cup \mathcal{C}^s, \varphi) \le K e^{n(P(\varphi)-\varepsilon)}, \tag{4.18}$$

$$\Lambda_n(\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s, \varphi) \le K e^{n(P(\varphi)-\varepsilon)} \tag{4.19}$$

for all n. From (4.19) we get $\Lambda_n(\varphi) \le \Lambda_n(\mathcal{C}^p\mathcal{G}\mathcal{C}^s, \varphi) + K e^{n(P(\varphi)-\varepsilon)}$, so it suffices to prove the upper bound in (4.14) for $\Lambda_n(\mathcal{C}^p\mathcal{G}\mathcal{C}^s, \varphi)$.

Write $a_j = \Lambda_j(\mathcal{G}, \varphi)e^{-jP(\varphi)}$, and observe that $a_j \le C$ by (4.17). Since every word $x \in (\mathcal{C}^p\mathcal{G}\mathcal{C}^s)_n$ can be decomposed as $x = uvw$ where $u \in \mathcal{C}^p$, $v \in \mathcal{G}$, and $w \in \mathcal{C}^s$, we have

$$\begin{aligned}
\Lambda_n(\mathcal{C}^p\mathcal{G}\mathcal{C}^s, \varphi) &\le \sum_{i+j+k=n} \Lambda_i(\mathcal{C}^p, \varphi)\Lambda_j(\mathcal{G}, \varphi)\Lambda_k(\mathcal{C}^s, \varphi) \\
&\le K^2 \sum_{i+j+k=n} e^{i(P(\varphi)-\varepsilon)} a_j e^{jP(\varphi)} e^{k(P(\varphi)-\varepsilon)} \\
&= K^2 e^{nP(\varphi)} \sum_{i+j+k=n} a_j e^{-(i+k)\varepsilon} \\
&= K^2 e^{nP(\varphi)} \sum_{m=0}^{n} a_{n-m}(m+1)e^{-m\varepsilon}.
\end{aligned} \tag{4.20}$$

Because $a_{n-m} \le C$ and $\sum_{m\ge 0}(m+1)e^{-m\varepsilon} < \infty$, this establishes the second half of (4.14).

Finally, we use (4.19) and (4.20) to show (4.15). Note that it suffices to produce $j \in (n-N, n]$ with $a_j \ge Q_3$. Using (4.19), (4.20) and the first half of (4.14), we have

$$e^{nP(\varphi)} - K e^{n(P(\varphi)-\varepsilon)} \le \Lambda_n(\mathcal{C}^p\mathcal{G}\mathcal{C}^s, \varphi) \le K^2 e^{nP(\varphi)} \sum_{m=0}^{n} a_{n-m}(m+1)e^{-m\varepsilon},$$

which yields $\frac{1}{2} \le 1 - K e^{-n\varepsilon} \le K^2 \sum_{m=0}^n a_{n-m}(m+1)e^{-m\varepsilon}$ whenever $n \ge n_0$, where n_0 is chosen such that $K e^{-n_0\varepsilon} < \frac{1}{2}$. Thus

$$\frac{1}{2}K^{-2} \le \sum_{m=0}^{N-1} a_{n-m}(m+1)e^{-m\varepsilon} + \sum_{m\ge N} Q_2(m+1)e^{-m\varepsilon},$$

using the inequality $a_{n-m} \le Q_2$. Now let N be large enough that $K' := \frac{1}{2}K^{-2} - Q_2 \sum_{m\ge N}(m+1)e^{-m\varepsilon} > 0$. Then

$$\sum_{m=0}^{N-1} a_{n-m}(m+1)e^{-m\varepsilon} \ge K',$$

and since $(m+1)e^{-m\varepsilon} \le 1/\varepsilon$ for every $m \ge 0$, we may put $Q_3 := \varepsilon K'/N$ and conclude that there is $0 \le m < N$ such that $a_{n-m} \ge Q_3$, which completes the proof of Lemma 4.5. □

Lemma 4.5 leads to the following important bound. Given $v \in \mathcal{L}_k$ and $1 \le i \le n-k$, consider

$$\mathcal{H}_n(v,i) = \{w \in \mathcal{L}_n \mid w_{[i,i+k)} = v\}, \tag{4.21}$$

the set of words where v appears starting in index i, but the entries of w before i and after $i+k$ are free to vary (this is the finite-length analogue of a cylinder set). We will mostly be interested in the case when $v \in \mathcal{G}$. In this case we have the following non-stationary version of the Gibbs property (for the measure-theoretic equivalent, see [CT13, §5.2] and §2.2.1).

Proposition 4.6. *There is $Q_4 > 0$ such that for every $1 \le i \le i+k \le n$, we have*

$$\begin{aligned} \Lambda_n(\mathcal{H}_n(v,i),\varphi) &\le Q_4 e^{(n-k)P(\varphi)+\hat\varphi(v)} \quad \textit{for every } v \in \mathcal{L}_k, \\ \Lambda_n(\mathcal{H}_n(v,i),\varphi) &\ge Q_4^{-1} e^{(n-k)P(\varphi)+\hat\varphi(v)} \textit{for every } v \in \mathcal{G}_k. \end{aligned} \tag{4.22}$$

Proof. For the upper bound, observe that (4.13) gives

$$\Lambda_n(\mathcal{H}_n(v,i),\varphi) \le \Lambda_i(\mathcal{L},\varphi) e^{\hat\varphi(v)} \Lambda_{n-(i+k)}(\mathcal{L},\varphi),$$

and using (4.14) gives $\Lambda_n(\mathcal{H}_n(v,i),\varphi) \le (Q_2)^2 e^{(n-k)P(\varphi)} e^{\hat\varphi(v)}$.

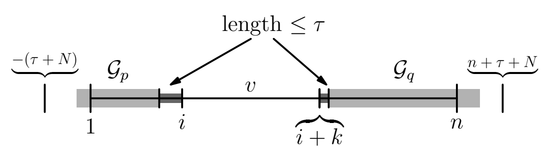


FIGURE 4.1. Estimating $\Lambda_n(\mathcal{H}_n(v,i))$.

For the lower bound, we use the usual specification argument, illustrated in Figure 4.1. By Lemma 4.5, there are $p, q \in \mathbb{N}$ such that

$$\begin{aligned} \Lambda_p(\mathcal{G},\varphi) &\ge Q_3 e^{pP(\varphi)} \text{ and } p \in [i, i+N], \\ \Lambda_q(\mathcal{G},\varphi) &\ge Q_3 e^{qP(\varphi)} \text{ and } i+k+q \in [n, n+N]. \end{aligned} \tag{4.23}$$

Let $w^1 \in \mathcal{G}_p$ and $w^2 \in \mathcal{G}_q$ be arbitrary. Then by **[I]** there are $u^1, u^2 \in \mathcal{L}$ with $|u^i| \le \tau$ such that $w^1u^1vu^2w^2 \in \mathcal{G}$. Note that $|w^1u^1| \in [i, i+N+\tau]$, and so by truncating at most $\tau + N$ symbols from the beginning and end of $w^1u^1vu^2w^2$, we obtain a word $T(w^1,w^2) \in \mathcal{H}_n(v,i)$ with the property that the first $i - |u^1|$ symbols of $T(w^1,w^2)$ match the last $i - |u^1|$ symbols of w^1, and similarly for the end of $T(w^1,w^2)$ and the beginning of w^2.

This defines a map $T\colon \mathcal{G}_p \times \mathcal{G}_q \to \mathcal{H}_n(v,i)$. Note that

$$\begin{aligned} \hat\varphi(T(w^1,w^2)) &\ge \hat\varphi(w^1u^1vu^2w^2) - (2N+2\tau)\|\varphi\| \\ &\ge \hat\varphi(w^1) + \hat\varphi(v) + \hat\varphi(w^2) - (2N+4\tau)\|\varphi\| - 2\,|\varphi|_d\,, \end{aligned} \tag{4.24}$$

Moreover, since the act of truncation removes at most $2(N+\tau)$ symbols from $w^1u^1vu^2w^2$, we see that each word in $\mathcal{H}_n(v,i)$ has at most $(\#A+1)^{2(N+\tau)}$

preimages under the map T. This yields the estimate

$$\begin{aligned}\Lambda_n(\mathcal{H}_n(v,i)) &\geq (\#A+1)^{-2(N+\tau)} \sum_{w^1\in\mathcal{G}_p} \sum_{w^2\in\mathcal{G}_q} e^{\hat\varphi(T(w^1,w^2))} \\ &\geq C \sum_{w^1\in\mathcal{G}_p} \sum_{w^2\in\mathcal{G}_q} e^{\hat\varphi(w^1)} e^{\hat\varphi(v)} e^{\hat\varphi(w^2)} \geq C(Q_3)^2 e^{\hat\varphi(v)} e^{-(n-k)P(\varphi)},\end{aligned}$$

where the first inequality uses the multiplicity bound, the second uses (4.24), and the third uses (4.23). ◻

We need one more counting estimate that we will use in the proof of Theorem 3.1, so now we assume that $\mathcal{F}$ satisfies $[\mathbf{I}_0]$ and $[\mathbf{II}]$, and let d be the gcd of the lengths $\{|w| \mid w \in \mathcal{F}\}$. Replacing σ with σ^d, we will assume without loss of generality that $d = 1$. Then we have the following estimate, which strengthens (4.15).

Lemma 4.7. *There is $Q_5 > 0$ such that $\Lambda_n(\mathcal{F},\varphi) \geq Q_5 e^{nP(\varphi)}$ for all sufficiently large n.*

Proof. Since $\gcd\{|w| \mid w \in \mathcal{F}\} = 1$, there is $m \in \mathbb{N}$ such that for every $n \geq m$ we have $n = \sum_{i=1}^k a_i |w^i|$ for some $a_i \in \mathbb{N}$ and $w^i \in \mathcal{F}$. Write $(w^i)^{a_i}$ for the word w^i repeated a^i times and note that by $[\mathbf{I}_0]$ we have $w := (w^1)^{a_1}\cdots(w^k)^{a_k} \in \mathcal{F}$, and $|w| = n$. Thus $\mathcal{F}_n$ is non-empty for every $n \geq m$.

Now by (4.15) there are $n_0, N \in \mathbb{N}$ and $Q_3 > 0$ such that for every $n \geq n_0 + m$ there is $j \in (n-m-N, n-m]$ with

$$\Lambda_j(\mathcal{F},\varphi) \geq Q_3 e^{jP(\varphi)} \geq Q_3 e^{-(m+N)P(\varphi)} e^{nP(\varphi)}. \tag{4.25}$$

Since $n - j \in [m, m+N)$, by the definition of m there is $w \in \mathcal{F}_{n-j}$; note that $|\hat\varphi(w)| \leq (n-j)\|\varphi\| \leq (m+N)\|\varphi\|$. Now we can use $[\mathbf{I}_0]$ to get

$$\Lambda_n(\mathcal{F},\varphi) \geq \sum_{v\in\mathcal{F}_j} e^{\hat\varphi(vw)} \geq e^{-\hat\varphi(w)-|\varphi|_d} \sum_{v\in\mathcal{F}_j} e^{\hat\varphi(v)} \geq e^{-\|\varphi\|(m+N)-|\varphi|_d} \Lambda_j(\mathcal{F},\varphi),$$

where the second inequality uses the first half of (4.12). Together with (4.25), this completes the proof of Lemma 4.7. ◻

5. Proof of Theorem 3.1

In this section we prove Theorem 3.1. Without loss of generality we assume that X is a two-sided shift space; if it is one-sided then we pass to the natural extension and define φ to depend only on non-negative coordinates. We will prove the conclusions of Theorem 3.1(B) for both the one-sided and two-sided shifts (see (5.1)).

Assume that we are given $\mathcal{F} \subset \mathcal{L}$ satisfying $[\mathbf{I}_0]$, so that $I = \mathcal{F} \setminus \mathcal{F}\mathcal{F}$ has $I^* = \mathcal{F} \subset \mathcal{L}$. Let Σ be the countable-state Markov shift constructed in §3.1, $T\colon \Sigma \to \Sigma$ the shift map, and $\pi\colon \Sigma \to X$ the one-block code given there. We will denote a typical element of Σ as $\mathbf{z} = \{\mathbf{z}_j\}_{j\in\mathbb{Z}}$, where each $\mathbf{z}_j$ is of the form (w,k) for some $w \in I$, $1 \leq k \leq |w|$.

We start by describing the overall structure of the proof. Let X^+ and Σ^+ be the one-sided versions of X and Σ respectively; that is, define $p\colon A^{\mathbb{Z}} \to A^{\mathbb{N}\cup\{0\}}$ by $p(\cdots x_{-1}x_0x_1\cdots) = x_0x_1\cdots$, let $X^+ = p(X)$, and then similarly define $\hat{p}\colon (A_I)^{\mathbb{Z}} \to (A_I)^{\mathbb{N}\cup\{0\}}$ and put $\Sigma^+ = \hat{p}(\Sigma)$. We write $\pi^+\colon \Sigma^+ \to X^+$ for the one-block code described above. Let $\mathcal{M}_T$ denote the space of T-invariant Borel probability measures, and similarly for $\mathcal{M}_\sigma$. Now we have the following commutative diagrams.

$$\begin{tikzcd} \Sigma \arrow[r, "\pi"] \arrow[d, "\hat{p}"] & X \arrow[d, "p"] \\ \Sigma^+ \arrow[r, "\pi^+"] & X^+ \end{tikzcd} \qquad \begin{tikzcd} \mathcal{M}_T(\Sigma) \arrow[r, "\pi_*"] \arrow[d, "\hat{p}_*"] & \mathcal{M}_\sigma(X) \arrow[d, "p_*"] \\ \mathcal{M}_T(\Sigma^+) \arrow[r, "\pi_*^+"] & \mathcal{M}_\sigma(X^+) \end{tikzcd} \tag{5.1}$$

The following facts regarding (5.1) are either immediate or well-known.

- The maps $\pi, \pi^+, p, \hat{p}$ commute with the shifts T, σ.
- The fact that $\pi(\Sigma) \subset X$ is a consequence of $[\mathbf{I}_0]$.
- Although $p, \hat{p}$ are not 1-1, the induced maps $p_*, \hat{p}_*$ on the space of invariant measures are 1-1, so that in particular the pullback of invariant measures from one-sided to two-sided shifts is well-defined; see [CT12, Proposition 2.1] for a proof.

We outline the remainder of the proof of Theorem 3.1 (recalling (3.4)), indicating where the conclusions (A.1)–(A.6) and (i)–(vi) are proved (note that (A.7) was proved in §3.3.3).

- **§5.1:** Condition $[\mathbf{III}^*]$ gives injectivity of π, for (A.1), and injectivity gives unique decipherability of I, for (A.2).
- **§5.2:** The potential φ induces a Hölder potential $\Phi = \varphi \circ \pi\colon \Sigma \to \mathbb{R}$, which is cohomologous to a Hölder potential $\Phi^+\colon \Sigma^+ \to \mathbb{R}$ via a standard procedure.[18] Unique decipherability lets the estimates from Lemma 4.7 be applied to the partition sums Z_n and Z_n^*, and then $[\mathbf{II}']$ gives strong positive recurrence, for (A.3); Sarig showed that this gives a unique equilibrium state m for (Σ^+, T, Φ^+) [Sar99].
- **§5.3:** Although π is not generally surjective, if $[\mathbf{II}']$ holds then every equilibrium state for (X, φ) gives full weight to $\pi(\Sigma)$, for (A.4). Moreover, (ii) and (iii) are satisfied for *some* equilibrium state of (X, φ) by Proposition 4.6 and arguments in [CT13], although uniqueness (i) does not yet follow.
- **§5.4:** If I is a uniquely decipherable generating set and $[\mathbf{II}']$ holds, then for every equilibrium state μ of (X, σ, φ), we have $\#\pi^{-1}(x) < \infty$ for μ-a.e. x, giving (A.5). Moreover, this property together with $\mu(\pi(\Sigma))$ guarantees that $\mu = \pi_*\nu$ for some invariant measure ν on Σ with $h(\nu) = h(\mu)$, which establishes (A.6) and also proves that μ is unique, giving (i) and hence (ii) and (iii) by §5.3.

[18] This process uses the Markov structure of Σ and hence cannot be used directly to get a potential $\varphi^+\colon X^+ \to \mathbb{R}$.

- **§5.5:** Strong positive recurrence gives (iv)–(vi) for (Σ^+, T, m) [Sar01, CS09], and since the unique equilibrium state μ for (X, φ) has $\mu = \pi_* m$, these conclusions hold for (X^+, σ, μ) as well. Standard arguments give the two-sided results from these.

5.1. $\mathcal{F}$-marking sets, injectivity of π, and unique decipherability.

Proposition 5.1. *If $\mathcal{F}$ satisfies* $[\mathbf{III}^*]$*, then $\pi\colon \Sigma \to X$ is injective. If π is injective, then I is uniquely decipherable (whether or not $\mathcal{F}$ satisfies* $[\mathbf{III}^*]$*).*

This section is devoted to the proof of Proposition 5.1, which is (A.1) and (A.2). First we set up some terminology to relate Σ and X. Say that a (finite or infinite) set $J \subset \mathbb{Z}$ is **$\mathcal{F}$-marking** for $x \in X$ if $x_{[i,j)} \in \mathcal{F}$ for all $i, j \in J$ with $i < j$. Call $i, j \in J$ **consecutive** if $k \notin J$ for all k between i and j. We record some immediate consequences of $[\mathbf{I}_0]$ as a lemma.

Lemma 5.2. *$J \subset \mathbb{Z}$ is $\mathcal{F}$-marking for x if and only if $x_{[i,j)} \in \mathcal{F}$ for all consecutive $i < j \in J$. In particular, the following are equivalent.*

(1) J is $\mathcal{F}$-marking.
(2) There are $a_k \to -\infty$ and $b_k \to \infty$ such that $J \cap [a_k, b_k]$ is $\mathcal{F}$-marking for every k.
(3) $J \cap [a, b]$ is $\mathcal{F}$-marking for every $a < b \in \mathbb{Z}$.

Say that $J \subset \mathbb{Z}$ is **bi-infinite** if $J \cap [0, \infty)$ and $J \cap (-\infty, 0]$ are both infinite. Say that $J \subset \mathbb{Z}$ is **maximally $\mathcal{F}$-marking** for x if there is no $\mathcal{F}$-marking set $J' \subset \mathbb{Z}$ with $J' \supsetneq J$. Recall that I is the collection of irreducible elements of $\mathcal{F}$. The following lemma collects properties of bi-infinite $\mathcal{F}$-marking sets, and relates these to multiplicity of the map π.

Lemma 5.3. (a) *A bi-infinite set $J \subset \mathbb{Z}$ is maximally $\mathcal{F}$-marking for x if and only if $x_{[i,j)} \in I$ for all consecutive $i < j \in J$.*
(b) *Given $\mathbf{z} \in \Sigma$, the set $J(\mathbf{z}) := \{j \mid \mathbf{z}_j = (w, 1)$ for some $w \in I\}$ is bi-infinite and maximally $\mathcal{F}$-marking for $\pi(\mathbf{z}) \in X$.*
(c) *If $J \subset \mathbb{Z}$ is bi-infinite and maximally $\mathcal{F}$-marking for $x \in X$, then there is exactly one $\mathbf{z} \in \Sigma$ such that $\pi(\mathbf{z}) = x$ and $J(\mathbf{z}) = J$.*
(d) *Given $x \in X$, we have $x \in \pi(\Sigma)$ iff there is a bi-infinite $\mathcal{F}$-marking set $J \subset \mathbb{Z}$ for x. There is a 1-1 correspondence between elements of $\pi^{-1}(x) \subset \Sigma$ and bi-infinite maximal $\mathcal{F}$-marking sets for x.*

Proof. We start with the forward direction of (a). Let J be maximally $\mathcal{F}$-marking. Then $x_{[i,j)} \in \mathcal{F}$ for all consecutive $i < j \in J$. Suppose for a contradiction that there are consecutive $i < j \in J$ for which $x_{[i,j)} \notin I$; then there is $k \in (i, j)$ such that $x_{[i,k)}, x_{[k,j)} \in \mathcal{F}$. By Lemma 5.2 this implies that $J \cup \{k\}$ is $\mathcal{F}$-marking, contradicting maximality.

For the other direction, suppose J is not maximal; then there is $k \in \mathbb{Z} \setminus J$ such that $J \cup \{k\}$ is $\mathcal{F}$-marking. Let $i < k < j$ be such that $i < j \in J$ are consecutive (here we use that J is bi-infinite). Then $x_{[i,j)} = x_{[i,k)} x_{[k,j)} \in \mathcal{F}\mathcal{F}$, so $x_{[i,j)} \notin I$.

Now we prove (b). To see that $J(\mathbf{z})$ is bi-infinite, pick any $j \notin J(\mathbf{z})$; then $\mathbf{z}_j = (w, k)$ for some $w \in I$ and $1 < k \leq |w|$, and the transition rules (3.3) guarantee that $\mathbf{z}_{j-k+1} = \mathbf{z}_{j+|w|-k+1} \in I \times \{1\}$.

For maximally $\mathcal{F}$-marking, by (a) it suffices to check that $x_{[i,j)} \in I$ whenever $i < j \in J(\mathbf{z})$ are consecutive. In this case we have $\mathbf{z}_i = (w, 1)$ and $\mathbf{z}_j = (v, 1)$ for $w, v \in I$, with $\mathbf{z}_\ell \notin I \times \{1\}$ for $i < \ell < j$, hence by (3.3) we have $|w| = j - i$, and $x_{[i,j)} = w$.

For (c), take $x \in X$ and $J \subset \mathbb{Z}$ as in the hypothesis and let $i < j \in J$ be consecutive and let $w = x_{[i,j)}$. Then $w \in I$ and any $\mathbf{z} \in \pi^{-1}(x)$ with $J(\mathbf{z}) = J$ must have $\mathbf{z}_{[i,j)} = (w, 1)(w, 2) \cdots (w, |w|)$. This gives uniqueness, and existence follows by observing that this condition defines a legal sequence $\mathbf{z} \in \Sigma$.

Every bi-infinite $\mathcal{F}$-marking set J is contained in a bi-infinite maximal $\mathcal{F}$-marking set: this follows by taking every pair of consecutive $i < j \in J$ and decomposing $x_{[i,j)}$ as a concatenation of elements of I. Adding the indices marking these decompositions to J gives a maximal $\mathcal{F}$-marking set $J' \supset J$. Together with (b) and (c), this proves (d). □

As (d) clarifies, Lemma 5.3(c) does not yet prove injectivity of π, since it is a priori possible that some $x \in X$ has multiple maximally $\mathcal{F}$-marking sets. To show injectivity of π it suffices to show that every $x \in X$ has at most one maximal bi-infinite $\mathcal{F}$-marking set. We accomplish this by showing that arbitrary unions of bi-infinite $\mathcal{F}$-marking sets are still bi-infinite and $\mathcal{F}$-marking. This is where we need **[III*]**.

Lemma 5.4. *Suppose $\mathcal{F}$ satisfies* **[III*]***, and let $\{J_\lambda\}_{\lambda \in \Lambda}$ be any collection of sets $J_\lambda \subset \mathbb{Z}$ such that each J_λ is $\mathcal{F}$-marking for x. Let $r < s \in \mathbb{Z}$ be such that $r > \min J_\lambda$ and $s < \max J_\lambda$ for all λ. Then $(\bigcup_\lambda J_\lambda) \cap [r, s]$ is $\mathcal{F}$-marking for x.*

Proof. Pick $j < k \in \bigcup_\lambda J_\lambda$ with $r \leq j < k \leq s$. Let λ, λ' be such that $j \in J_\lambda$ and $k \in J_{\lambda'}$. Because $\min J_{\lambda'} \leq r$ there is $i \in J_{\lambda'}$ with $i \leq j$; similarly, $\max J_\lambda \geq s$ implies that there is $\ell \in J_\lambda$ with $\ell \geq k$. Thus $i \leq j < k \leq \ell$ are such that $x_{[i,k)}, x_{[j,\ell)} \in \mathcal{F}$. (See Figure 5.1.)

Moreover, choosing $a \in J_\lambda$ with $a < r$ and $b \in J_{\lambda'}$ with $b > s$, we have $x_{[a,j)}, x_{[k,b)} \in \mathcal{F}$, and it follows from Condition **[III*]** that $x_{[j,k)} \in \mathcal{F}$. This holds for all $j, k \in (\bigcup_\lambda J_\lambda) \cap [r, s]$, so we are done. □

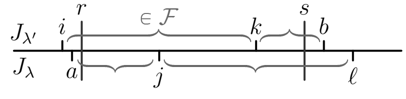


FIGURE 5.1. The union of $\mathcal{F}$-marking sets is $\mathcal{F}$-marking.

When the sets J_λ in Lemma 5.4 are all bi-infinite, we can take $r \to -\infty$ and $s \to \infty$ in Lemma 5.2. We conclude that if $\mathcal{F}$ has **[III*]** and $\{J_\lambda\}_{\lambda \in \Lambda}$ is

any collection of bi-infinite $\mathcal{F}$-marking sets $J_\lambda \subset \mathbb{Z}$ for $x \in X$, then $\bigcup_\lambda J_\lambda$ is bi-infinite and $\mathcal{F}$-marking for x. Now we can prove injectivity of π as follows: given $x \in \pi(\Sigma)$, let $\{J_\lambda\}_{\lambda\in\Lambda}$ be the collection of *all* bi-infinite $\mathcal{F}$-marking sets for x. This collection is non-empty by Lemma 5.3(b). Let $J = \bigcup_\lambda J_\lambda$. We have just shown that J is bi-infinite and $\mathcal{F}$-marking for x. Moreover, if J' is any bi-infinite $\mathcal{F}$-marking set, we have $J' \subset J$ by construction, so J is maximal, and it is the only bi-infinite maximal $\mathcal{F}$-marking set. Thus π is 1-1, which completes the proof of the first part of Proposition 5.1.

For the second part of Proposition 5.1, we prove the contrapositive. Suppose I is *not* uniquely decipherable, and let $v^1, \dots, v^k, w^1, \dots, w^\ell \in I$ be such that $v^1 \cdots v^k = w^1 \cdots w^\ell$ but there is i such that $v^i \neq w^i$. Then $(v^1, \dots, v^k)$ and $(w^1, \dots, w^\ell)$ correspond to distinct periodic points $\mathbf{y}, \mathbf{z} \in \Sigma$ such that $\pi(\mathbf{y}) = \pi(\mathbf{z})$, so π is not injective.

5.2. **One-sided shifts and strong positive recurrence.** From (2.4) we have $|\varphi(x) - \varphi(y)| \leq |\varphi|_\beta e^{-\beta n}$ whenever $x, y \in X$ have $x_k = y_k$ for all $|k| \leq n$. Define $\Phi\colon \Sigma \to \mathbb{R}$ by $\Phi = \varphi \circ \pi$; then for every $\mathbf{z}, \mathbf{z}' \in \Sigma$ with $\mathbf{z}_k = \mathbf{z}'_k$ for all $|k| \leq n$, we have $\pi(\mathbf{z})_k = \pi(\mathbf{z}')_k$ for all $|k| \leq n$ (since π is a one-block code) and hence $|\Phi(\mathbf{z}) - \Phi(\mathbf{z}')| \leq |\varphi|_\beta e^{-\beta n}$. In particular, $\Phi \in C^h(\Sigma)$. We need the following result, which goes back (in a different context) to Sinai; for a proof, see [Bow75, Lemma 1.6] (the statement there is for subshifts of finite type, but the proof goes through in our setting as well, see [Sar11, Lemma 3.3]).

Lemma 5.5. *Given $\psi \in C^h(\Sigma)$, there is a bounded function $u \in C^h(\Sigma)$ such that the function $\psi^+ := \psi - u + u \circ T \in C^h(\Sigma)$ only depends on non-negative coordinates; that is, $\psi^+(\mathbf{z}) = \psi^+(\mathbf{z}')$ whenever $\mathbf{z}_k = \mathbf{z}'_k$ for all $k \geq 0$. The maps $\psi \mapsto u$ and $\psi \mapsto \psi^+$ are linear.*

The map ψ^+ can be considered as a function $\Sigma^+ \to \mathbb{R}$, and is Hölder continuous with the same constant and exponent as $\psi^+\colon \Sigma \to \mathbb{R}$. Finally, for any $\mathbf{z} \in \Sigma$ we have

$$\left| S_n\psi^+(\hat{p}(\mathbf{z})) - S_n\psi(\mathbf{z}) \right| \leq 2\|u\|. \tag{5.2}$$

Applying Lemma 5.5 to the function $\Phi = \varphi \circ \pi\colon \Sigma \to \mathbb{R}$, we obtain a Hölder function $u\colon \Sigma \to \mathbb{R}$ such that $\Phi^+ = \Phi - u + u \circ T$ depends only on the non-negative coordinates, and thus may be considered as a Hölder continuous function on Σ^+. Fix $v \in I$ and let $\mathbf{a} = (v, 1) \in A_I$; recall the definition of $Z_n(\Phi^+, \mathbf{a})$ and $Z_n^*(\Phi^+, \mathbf{a})$ in (2.6)–(2.7). To prove (A.3), we first assume that I is uniquely decipherable and relate Z_n, Z_n^* to partition sums on X; then we assume $P(I, \varphi) < P(\varphi)$ and use this to prove (2.10).

Lemma 5.6. *There is $Q_6 > 0$ such that if I is uniquely decipherable, then for every n we have*

$$e^{-Q_6}\Lambda_{n-|v|}(\mathcal{F}, \varphi) \leq Z_n(\Phi^+, \mathbf{a}) \leq e^{Q_6}\Lambda_{n-|v|}(\mathcal{F}, \varphi),$$
$$e^{-Q_6}\Lambda_{n-|v|}((I \setminus \{v\})^*, \varphi) \leq Z_n^*(\Phi^+, \mathbf{a}) \leq e^{Q_6}\Lambda_{n-|v|}((I \setminus \{v\})^*, \varphi).$$

Proof. Given $w \in I$, let $\tilde{w} = (w,1)\cdots(w,|w|) \in (A_I)^*$. Given $w \in \mathcal{F} = I^*$, let $w^1, \dots, w^k \in I$ be such that $w = w^1 \cdots w^k$; note that this k-tuple is unique because I is uniquely decipherable. Let $\tau(w) \in \Sigma$ be the periodic point of Σ obtained by repeating the word $\tilde{v}\widetilde{w^1}\cdots\widetilde{w^k}$. Then every periodic $\mathbf{z} \in \Sigma$ based at the vertex $\mathbf{a} = (v,1)$ is of the form $\tau(w)$ for some $w \in \mathcal{F}$ (note that the empty word is in $\mathcal{F}$), and so

$$
\begin{aligned}
Z_n(\Phi^+, \mathbf{a}) &= \sum_{w \in \mathcal{F}_{n-|v|}} e^{S_n \Phi^+(\hat{p}\circ\tau(w))}, \\
Z_n^*(\Phi^+, \mathbf{a}) &= \sum_{w \in ((I\setminus\{v\})^*)_{n-|v|}} e^{S_n \Phi^+(\hat{p}\circ\tau(w))}.
\end{aligned}
\tag{5.3}
$$

As in (4.11) we have $|S_n\varphi(x) - S_n\varphi(y)| \leq |\varphi|_d < \infty$ for all $x, y \in [w], w \in \mathcal{L}_n$. Together with (5.2), we see that for every $w \in \mathcal{F}_n$ we have

$$\left| S_n\Phi^+(\hat{p}(\tau(w))) - \hat{\varphi}(w) \right| \leq |\varphi|_d + 2\|u\| + |v|\,\|\varphi\|.$$

Along with (5.3), this completes the proof of Lemma 5.6. □

By Lemmas 4.5 and 4.7 we have

$$Q_5 e^{nP(\varphi)} \leq \Lambda_n(\mathcal{F}, \varphi) \leq Q_2 e^{nP(\varphi)}$$

for all sufficiently large n, and so

$$Q_5 e^{-Q_6} e^{(n-|v|)P(\varphi)} \leq Z_n(\Phi^+, \mathbf{a}) \leq Q_2 e^{Q_6} e^{(n-|v|)P(\varphi)}. \tag{5.4}$$

This implies that $P(X,\varphi) = P_G(\Sigma, \Phi^+)$ (see (2.8)), and that Φ^+ is positive recurrent. In fact, if $P(I,\varphi) < P(\varphi)$, then Φ^+ is strongly positive recurrent. This follows from (2.10), Lemma 5.6, and the following result.

Lemma 5.7. *If $I \subset \mathcal{L}$ is such that $I^* \subset \mathcal{L}$ and $P(I,\varphi) < P(I^*,\varphi)$, then for every $\hat{I} \subsetneq I$ we have $P(\hat{I}^*,\varphi) < P(I^*,\varphi)$.*

Proof. Given $w \in I^*$, let $0 = j_0 < j_1 < \cdots < j_\ell < j_{\ell+1} = |w|$ be such that $w_{(j_i, j_{i+1}]} \in I$ for every $0 \leq i \leq \ell$.[19] Given $\ell < n \in \mathbb{N}$, let $\mathbb{J}_\ell = \{J \subset [1,n) \mid \#J = \ell\}$; for each $J \in \mathbb{J}_\ell$, let

$$\mathcal{X}_n(J) = \{w \in (I^*)_n \mid \ell(w) = \ell \text{ and } \{j_i(w)\}_{i=1}^{\ell} = J\}.$$

Given $\delta > 0$, let $\mathcal{R}_\delta = \{w \in I^* \mid \#\ell(w) \leq \delta\,|w|\}$; we will prove Lemma 5.7 by showing that for sufficiently small values of δ, we have

$$P(I^* \setminus \mathcal{R}_\delta, \varphi) < P(I^*, \varphi), \tag{5.5}$$

$$P(\hat{I}^* \cap \mathcal{R}_\delta, \varphi) < P(I^*, \varphi), \tag{5.6}$$

[19]For purposes of this lemma we do not need unique decipherability, so the sequence j_i may not be uniquely defined; we can select any sequence that does the job. Note that the application of the lemma to obtain strong positive recurrence does require unique decipherability in order to use Lemma 5.6.

and then applying (2.3) to $\hat{I}^* \subset (\hat{I}^* \cap \mathcal{R}_\delta) \cup (I^* \setminus \mathcal{R}_\delta)$. To prove (5.5), we start by writing $(I^* \setminus \mathcal{R}_\delta)_n = \bigcup_{\ell=0}^{\lfloor \delta n \rfloor} \bigcup_{J \in \mathbb{J}_\ell} \mathcal{X}_n(J)$, so that

$$\Lambda_n(I^* \setminus \mathcal{R}_\delta) \leq \sum_{\ell=0}^{\lfloor \delta n \rfloor} \sum_{J \in \mathbb{J}_\ell} \Lambda_n(\mathcal{X}_n(J), \varphi). \tag{5.7}$$

To get (5.5), we will get upper bounds on $\#\mathbb{J}_\ell$ and on $\Lambda_n(\mathcal{X}_n(J), \varphi)$. For the first of these, we observe that $\#\mathbb{J}_\ell \leq \binom{n}{\ell}$ and use the following useful result.

Lemma 5.8. *Given $\delta \in (0,1)$, write $h(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta)$ for the standard entropy function. Then for every $n \in \mathbb{N}$ and $0 \leq \ell \leq n$, we have $\binom{n}{\ell} \leq (n+1)e^{h(\frac{\ell}{n})n+1}$.*

Proof. We use the bound $k \log k - k + 1 \leq \log(k!) \leq k \log k - k + 1 + \log(k+1)$, which can be obtained by integrating $\log t$ over $[1,k]$ (for the lower bound) and over $[1, k+1]$ (for the upper). This gives

$$\begin{aligned} \log \tbinom{n}{\ell} &= \log(n!) - \log(\ell!) - \log(n-\ell)! \\ &\leq n \log n + 1 + \log(n+1) - \ell \log \ell - (n-\ell)\log(n-\ell) \\ &= h(\tfrac{\ell}{n})n + 1 + \log(n+1), \end{aligned} \tag{5.8}$$

which proves the lemma. □

Given $\delta \in (0, \frac{1}{2})$ and $0 \leq \ell \leq \delta n$, we conclude from Lemma 5.8 that $\#\mathbb{J}_\ell \leq (n+1)e^{h(\delta)n+1}$. To bound $\Lambda_n(\mathcal{X}_n(J), \varphi)$, fix $\varepsilon > 0$ such that $P(I, \varphi) < P(I^*, \varphi) - 2\varepsilon$; then there is K such that $\Lambda_j(I, \varphi) \leq K e^{j(P(I^*, \varphi) - \varepsilon)}$ for all j, and so

$$\Lambda_n(\mathcal{X}_n(J), \varphi) \leq \prod_{i=0}^{\ell} \Lambda_{j_{i+1} - j_i}(I, \varphi) \leq K^{\ell+1} e^{n(P(I^*, \varphi) - \varepsilon)}.$$

Together with (5.7) and the bound on $\#\mathbb{J}_\ell$, this gives

$$\Lambda_n(I^* \setminus \mathcal{R}_\delta) \leq (n+1)e^{h(\delta)n+1} K^{\delta n+1} e^{n(P(I^*, \varphi) - \varepsilon)},$$

and (5.5) follows by taking δ small enough that $h(\delta) + \delta \log(K) < \varepsilon$.

To prove (5.6), fix $v \in I \setminus \hat{I}$ and consider for each $0 \leq k \leq \delta n$ the collection

$$\mathcal{A}_n^k = \{ w \in (\mathcal{R}_\delta)_n \mid w_{(j_{i-1}(w), j_i(w)]} \neq v \text{ for all } 1 \leq i \leq k \};$$

this contains all words $w \in (I^*)_n$ such that $w = u^1 \cdots u^\ell$ for $u^i \in I$ and $\ell > \delta n$, and moreover $u^i \neq v$ for any $1 \leq i \leq k$. In particular, we have $\mathcal{A}_n^{k+1} \subset \mathcal{A}_n^k$ and $(\hat{I}^* \cap \mathcal{R}_\delta) \subset \mathcal{A}_n^{\lfloor \delta n \rfloor}$, so we can estimate $\Lambda_n(\hat{I}^* \cap \mathcal{R}_\delta, \varphi)$ by estimating $\Lambda_n(\mathcal{A}_n^{k+|v|}, \varphi)/\Lambda_n(\mathcal{A}_n^k, \varphi)$.

To this end, let $d = \gcd\{|u| \mid u \in \hat{I}\}$, and let $\bar{v} = vv \cdots v$, where v is repeated d times so that $|\bar{v}| = d\,|v| =: m$. Let N be such that $(\hat{I}^*)_{dn} \neq \emptyset$

for every $dn > N$. Now given $n \in \mathbb{N}$ and $1 \le k \le \delta n - (N+m)$, we have $u\bar{v}w \in \mathcal{A}_n^k \setminus \mathcal{A}_n^{k+|v|}$ for every $u \in (\hat{I}^k)_{\le n}$ and $w \in (I^*)_{n-|u\bar{v}|}$, so

$$\begin{aligned}\Lambda_n(\mathcal{A}_n^k \setminus \mathcal{A}_n^{k+|v|}, \varphi) &\ge \sum_{u \in (\hat{I}^k)_{\le n}} \sum_{w \in (I^*)_{n-|u\bar{v}|}} e^{\hat{\varphi}(u)+\hat{\varphi}(\bar{v})+\hat{\varphi}(w)-2|\varphi|_d} \\ &= e^{\hat{\varphi}(\bar{v})-2|\varphi|_d} \sum_{u \in (\hat{I}^k)_{\le n}} e^{\hat{\varphi}(u)} \Lambda_{n-|u\bar{v}|}(I^*, \varphi).\end{aligned}$$

From Lemma 4.7 and the divisibility condition above, we have

$$\Lambda_{n-|u\bar{v}|}(I^*, \varphi) \ge Q_5^{-1} e^{(n-|u|-m)P(\varphi)} \ge Q_5^{-2} e^{-mP(\varphi)} \Lambda_{n-|u|}(I^*, \varphi)$$

for all sufficiently large n, and we conclude that

$$\begin{aligned}\Lambda_n(\mathcal{A}_n^k \setminus \mathcal{A}_n^{k+|v|}, \varphi) &\ge Q_5^{-2} e^{\hat{\varphi}(\bar{v})-2|\varphi|_d-mP(\varphi)} \sum_{u \in (\hat{I}^k)_{\le n}} \sum_{w \in (I^*)_{n-|u|}} e^{\hat{\varphi}(u)} e^{\hat{\varphi}(w)} \\ &\ge \gamma \Lambda_n(\mathcal{A}_n^k, \varphi),\end{aligned}$$

where $\gamma = Q_5^{-2} e^{\hat{\varphi}(\bar{v})-2|\varphi|_d-mP(\varphi)}$. In particular, we get $\Lambda_n(\mathcal{A}_n^{k+|v|}, \varphi) \le (1-\gamma)\Lambda_n(\mathcal{A}_n^k, \varphi)$. Using the fact that $\mathcal{A}_n^0 = (\mathcal{R}_\delta)_n \subset (I^*)_n$ and $(\hat{I}^* \cap \mathcal{R}_\delta)_n \subset \mathcal{A}_n^{\lfloor \delta n \rfloor}$, we get

$$\Lambda_n(\hat{I}^*, \varphi) \le \Lambda_n(\mathcal{A}_n^{\lfloor \delta n \rfloor}, \varphi) \le (1-\gamma)^{\lfloor \frac{\lfloor \delta n \rfloor - (N+m)}{|v|} \rfloor} \Lambda_n(I^*, \varphi).$$

Sending $n \to \infty$ gives

$$P(\hat{I}^* \cap \mathcal{R}_\delta, \varphi) \le P(I^*, \varphi) + \tfrac{\delta}{|v|} \log(1-\gamma) < P(I^*, \varphi),$$

which proves (5.6) and completes the proof of Lemma 5.7. □

By unique decipherability, Lemma 5.6 gives $P_G(0) = P(\mathcal{F}, 0) = h(\mathcal{F}) \le h(X)$, hence every ergodic T-invariant probability measure m on Σ^+ has $h(m) \le P_G(0) \le h(X)$. Moreover, $\int \Phi^+ \, dm < \infty$ for every m since Φ^+ is bounded. Because Φ^+ is positive recurrent on Σ^+, [Sar99, Theorem 4] gives the existence of a σ-finite measure ν on Σ^+ and a function $h > 0$ such that $L^*\nu = e^{P(\varphi)}\nu$, $Lh = e^{P(\varphi)}h$, and $\nu(h) = 1$, where L is the **Ruelle–Perron–Frobenius operator** associated to the potential Φ^+. The measure m defined by $dm = h \, d\nu$ is a T-invariant probability measure. By [Sar99, Theorem 7] and the remarks above, it can be characterised as the only T-invariant probability measure such that[20]

$$h(m) + \int \Phi^+ \, dm = P_G(\Phi^+) = P(\varphi). \tag{5.9}$$

(See also [BS03, Theorems 1.1 and 1.2], and [Sar15, Theorem 5.5].)

[20]In the setting of [Sar99], some care must be taken to deal with the possibility that we may have $h(m) = \infty$ and/or $\int \psi \, dm = -\infty$. Because X has finite topological entropy and φ is bounded, this is not a problem for us.

5.3. **Equilibrium states charge the tower.** Now we show that if **[II′]** holds, then any equilibrium state μ for (X,φ) gives full weight to $\pi(\Sigma)\subset X$. It suffices to consider the case when μ is ergodic. We use the criterion from Lemma 5.3(d): given $x\in X$, we have $x\in\pi(\Sigma)$ if and only if x has a bi-infinite $\mathcal{F}$-marking set $J\subset\mathbb{Z}$.

We want to construct such a set as follows: given $x\in X$, we consider the words $x_{[-n,n]}$. Unless $x_{[-n,n]}\in\mathcal{L}\setminus\mathcal{E}^p\mathcal{F}\mathcal{E}^s$ (which should be rare since this set has small pressure), there are $a<b\in[-n,n]$ such that $x_{[-n,a)}\in\mathcal{E}^p$, $x_{[a,b)}\in\mathcal{F}$, and $x_{[b,n]}\in\mathcal{E}^s$. Since $\mathcal{F}=I^*$, the word $x_{[a,b)}$ can be decomposed as a composition of words in I; the positions where these words start and end gives a $\mathcal{F}$-marking set $J\subset[a,b]$.

In order to use the sets J to produce a bi-infinite $\mathcal{F}$-marking set for x, we need some information about which words $x_{[i,j)}$ can be contained in a single element of I, $\mathcal{E}^p$, or $\mathcal{E}^s$. Let $\mathcal{E}'=I\cup\mathcal{E}^p\cup\mathcal{E}^s\cup(\mathcal{L}\setminus\mathcal{E}^p\mathcal{F}\mathcal{E}^s)\subset\mathcal{L}$. By **[II′]**, we have $P(\mathcal{E}',\varphi)<P(\varphi)$. We say that $R\subset\mathbb{Z}$ is **$\mathcal{E}'$-restricting for** $x\in X$ if for every $i<j\in\mathbb{Z}$ with $x_{[i,j)}\in\mathcal{E}'$, the interval $[i,j]$ contains at most one element of R.

First we prove that to check $x\in\pi(\Sigma)$ it suffices to check existence of a bi-infinite $\mathcal{E}'$-restricting set; then we show that μ-a.e. $x\in X$ has such a set.

Lemma 5.9. *If $x\in X$ has a bi-infinite $\mathcal{E}'$-restricting set $R\subset\mathbb{Z}$, then it has a bi-infinite $\mathcal{F}$-marking set J, and hence $x\in\pi(\Sigma)$ by Lemma 5.3*(d).

Proof. Enumerate R as $R=\{r_n\}_{n\in\mathbb{Z}}$ where r_n is increasing (see Figure 5.2). Given $n\in\mathbb{N}$, note that $x_{[r_{-n},r_n)}\in\mathcal{E}^p\mathcal{F}\mathcal{E}^s$ since otherwise we would have a word in $\mathcal{L}\setminus\mathcal{E}^p\mathcal{F}\mathcal{E}^s\subset\mathcal{E}'$ that crosses more than one index in R. Thus we can apply the decomposition $\mathcal{E}^p\mathcal{F}\mathcal{E}^s$ to $x_{[r_{-n},r_n)}$ to get $j'\le j''\in[r_{-n},r_n)$ such that

$$x_{[r_{-n},j')}\in\mathcal{E}^p,\quad x_{[j',j'')}\in\mathcal{F},\quad x_{[j'',r_n)}\in\mathcal{E}^s.$$

Now use the decomposition $\mathcal{F}=I^*=\bigcup_{k\in\mathbb{N}}I^k$ to get an increasing sequence $\{j_i^n\}_{i=1}^k$ such that $j_1^n=j'$, $j_k^n=j''$, and $x_{[j_i^n,j_{i+1}^n)}\in I$ for every $1\le i<k$.

Put $j_0^n=r_{-n}$ and $j_{k+1}^n=r_n$. Then $x_{[j_i^n,j_{i+1}^n)}\in\mathcal{E}'$ for every $0\le i\le k$. Let $J_n=\{j_i^n\}_{i=1}^k$ and note that J_n is $\mathcal{F}$-marking for x. Since R is $\mathcal{E}'$-restricting, for every $\ell\in[-n,n)$ the interval $[r_\ell,r_{\ell+1})$ contains at least one element of J_n.

By the previous paragraph, there is $s_0\in[r_0,r_1)$ such that the set $\mathbf{N}_0:=\{n\in\mathbb{N}\mid s_0\in J_n\}$ is infinite. Similarly, there are $s_1\in[r_1,r_2)$ and $s_{-1}\in[r_{-1},r_0)$ such that $\mathbf{N}_1:=\{n\in\mathbf{N}_0\mid s_1,s_{-1}\in J_n\}$ is infinite. Continuing in this manner, we choose for each $\ell\in\mathbb{N}$ two indices $s_\ell\in[r_\ell,r_{\ell+1})$ and $s_{-\ell}\in[r_{-\ell},r_{-\ell+1})$ such that $\mathbf{N}_\ell=\{n\in\mathbf{N}_{\ell-1}\mid s_\ell,s_{-\ell}\in J_n\}$ is infinite. It follows from the definition of J_n that $x_{[s_\ell,s_{\ell+1})}\in\mathcal{F}$ for every $\ell\in\mathbb{Z}$, so $J:=\{s_\ell\}_{\ell\in\mathbb{Z}}$ is a bi-infinite $\mathcal{F}$-marking set for x. □

Lemma 5.10. *Let μ be any equilibrium state for (X,φ). Then for μ-a.e. $x\in X$, there is $n=n(x)$ such that for all $k\ge n$ and all $\ell\in[0,k]$ we have*

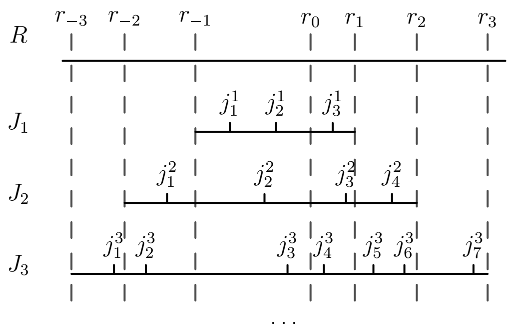


FIGURE 5.2. Constructing a bi-infinite $\mathcal{F}$-marking set.

$x_{[-\ell,k-\ell)} \notin \mathcal{E}'$. *In particular, the measure of the following sets decays to 0 as* $n \to \infty$:[21]

$$B_n := \{x \in X \mid x_{[-\ell,k-\ell)} \in \mathcal{E}' \textit{ for some } k \geq n \textit{ and } \ell \in [0,k]\} \tag{5.10}$$

Proof. Without loss of generality we may assume that μ is ergodic. Fix $\varepsilon > 0$ such that $P(\varphi) - 5\varepsilon > P(\mathcal{E}',\varphi)$. By Theorems 4.1 and 4.2, for μ-a.e. $x \in X$ there is $N_x \in \mathbb{N}$ such that for all $n \geq N_x$ and $\ell \in [0,n]$ we have

$$\begin{aligned} \mu[x_{[-\ell,-\ell+n)}] &\leq e^{-nh(\mu)+n\varepsilon}, \\ S_n\varphi(\sigma^{-\ell}x) &\geq n\left(\int \varphi\, d\mu - \varepsilon\right), \\ n\varepsilon &\geq |\varphi|_d\,. \end{aligned} \tag{5.11}$$

Let $A_n = \{x \in X \mid N_x \leq n\}$; then $\mu(A_n) \to 1$ as $n \to \infty$. If $k \geq n$ and $w \in \mathcal{L}_k$ is such that $\sigma^\ell[w] \cap A_n \neq \emptyset$ for some $0 \leq \ell \leq k$, then we can choose x in the intersection, so that $x_{[-\ell,-\ell+k)} = w$ and $k \geq N_x$; then (5.11) gives

$$\begin{aligned} \mu[w] &\leq e^{-kh(\mu)+k\varepsilon}, \\ \hat{\varphi}(w) &\geq (S_k\varphi(\sigma^{-\ell}x)) - |\varphi|_d \geq k\left(\int \varphi\, d\mu - 2\varepsilon\right), \end{aligned}$$

so we get

$$\mu[w]e^{-\hat{\varphi}(w)} \leq e^{-k(h(\mu)+\int \varphi\, d\mu - 3\varepsilon)} = e^{-k(P(\varphi)-3\varepsilon)} \leq e^{-kP(\mathcal{E}',\varphi)}e^{-2k\varepsilon}, \tag{5.12}$$

and summing over all long cylinders that intersect $A_n \cap B_n$ gives

$$\begin{aligned} &\mu(A_n \cap B_n) \leq \sum_{k\geq n} \sum_{w\in\mathcal{E}'_k} \sum_{\ell=0}^{k} \mu(\sigma^\ell[w] \cap A_n) \\ &\leq \sum_{k\geq n}(k+1) \sum_{w\in\mathcal{E}'_k} e^{-kP(\mathcal{E}',\varphi)}e^{-2k\varepsilon}e^{\hat{\varphi}(w)} = \sum_{k\geq n}(k+1)e^{-kP(\mathcal{E}',\varphi)}e^{-2k\varepsilon}\Lambda_k(\mathcal{E}',\varphi). \end{aligned}$$

[21] If μ is known to satisfy the upper Gibbs bound, then a simple computation shows that in fact $\mu(B_n)$ decays exponentially, but we will not need this.

Choose C such that $\Lambda_k(\mathcal{E}',\varphi) \le Ce^{k(P(\mathcal{E}',\varphi)+\varepsilon)}$ for all k. Then

$$\mu(A_n \cap B_n) \le C \sum_{k\ge n} (k+1)e^{-k\varepsilon} \to 0 \text{ as } n \to \infty,$$

so $\mu(B_n) \le \mu(X \setminus A_n) + \mu(A_n \cap B_n) \to 0$ as $n \to \infty$. In particular, we get $\mu(\bigcap_{n\in\mathbb{N}} B_n) = 0$, and so for μ-a.e. $x \in X$ there is $n > 0$ such that $x \notin B_n$, which proves Lemma 5.10. ◻

Lemma 5.11. *Let μ be any equilibrium state for (X,φ). Then μ-a.e. $x \in X$ has a bi-infinite $\mathcal{E}'$-restricting set. In particular, $\mu(\pi(\Sigma)) = 1$.*

Proof. Let E be the set of points satisfying the conclusion of Lemma 5.10; that is, for every $x \in E$ there is $n(x)$ such that for all $k \ge n(x)$ and all $\ell \in [0,k]$ we have $x_{[-\ell,k-\ell)} \notin \mathcal{E}'$. By Lemma 5.10 we have $\mu(X\setminus E) = 0$, hence $\mu(\sigma^{-m}(X\setminus E)) = 0$ for every $m \in \mathbb{Z}$, and we conclude that $E' := \bigcap_{m\in\mathbb{Z}} \sigma^{-m}E$ has full μ-measure. For every $x \in E'$ and every $m \in \mathbb{Z}$ there is $n(m) \in \mathbb{N}$ such that for all $a \le m \le b$ with $b - a \ge n(m)$, we have $x_{[a,b)} \notin \mathcal{E}'$.

Given $x \in E'$, define a bi-infinite sequence $r_j \in \mathbb{Z}$ by

(1) $r_0 = 0$;
(2) $r_{j+1} = r_j + n(r_j)$ for $j \ge 0$;
(3) $r_{j-1} = r_j - n(r_j)$ for $j \le 0$.

Let $R = \{r_j\}_{j\in\mathbb{Z}}$, and note that R is bi-infinite. We claim that R is $\mathcal{E}'$-restricting for x. Note that by the construction of R, we have $r_{j+1} - r_j \ge \min(n(r_j), n(r_{j+1}))$ for every $j \in \mathbb{Z}$. Thus if $a < b \in \mathbb{Z}$ are such that $a \le r_j$ and $b \ge r_{j+1}$, we either have $b-a \ge n(r_j)$ or $b-a \ge n(r_{j+1})$. It follows from the definition of n that $x_{[a,b)} \notin \mathcal{E}'$, since $r_j, r_{j+1} \in [a,b]$, and we conclude that R is $\mathcal{E}'$-restricting for x. ◻

Lemma 5.11 completes the proof of (A.4). Before proving liftability, we demonstrate that (ii) holds for some equilibrium state μ of (X,φ), and that every limiting measure of the periodic orbit measures is an equilibrium state; once uniqueness is established, these two results will give (ii) and (iii) from Theorem 1.1.

Recall a standard construction of an equilibrium state for (X,σ,φ): for each $w \in \mathcal{L}$, let $x(w) \in [w]$ be the point that maximises $S_{|w|}\varphi$; then consider the measures defined by

$$\nu_n = \frac{1}{\Lambda_n(\mathcal{L},\varphi)} \sum_{w\in\mathcal{L}_n} e^{\hat\varphi(w)}\delta_{x(w)}, \qquad \mu_n = \frac{1}{n}\sum_{k=0}^{n-1} \sigma_*^k \nu_n, \tag{5.13}$$

where δ_x is the point mass at x, and let μ be a weak* limit point of the sequence μ_n. It is shown in [Wal82, Theorem 9.10] that $h(\mu) + \int \varphi\,d\mu = P(\varphi)$, so μ is an equilibrium state. We can obtain the (non-uniform) Gibbs property for μ using Proposition 4.6: the following result mimics [CT13, Proposition 5.5 and Lemma 5.6], which are formally mildly weaker.

Proposition 5.12. *μ has the Gibbs property (2.11) for φ with respect to $\mathcal{F}$.*

Proof. Fix $n \in \mathbb{N}$ and $w \in \mathcal{L}_n$. Then for $m > n$ and $1 \leq k < m - n$, we have

$$\nu_m(\sigma^{-k}[w]) = \frac{\Lambda_m(\mathcal{H}_m(w,k),\varphi)}{\Lambda_m(\mathcal{L},\varphi)} \leq \frac{Q_4 e^{(m-|w|)P(\varphi)+\hat{\varphi}(w)}}{e^{mP(\varphi)}},$$

where the inequality uses Proposition 4.6 for the numerator and Lemma 4.5 (specifically the first inequality of (4.14)) for the denominator. Sending $m \to \infty$ gives the upper Gibbs bound in (2.11). To prove the lower Gibbs bound in (2.11), we observe that when $w \in \mathcal{F}$, we have

$$\nu_m(\sigma^{-k}[w]) = \frac{\Lambda_m(\mathcal{H}_m(w,k),\varphi)}{\Lambda_m(\mathcal{L},\varphi)} \geq \frac{Q_5 e^{(m-|w|)P(\varphi)+\hat{\varphi}(w)}}{Q_2 e^{mP(\varphi)}}$$

for all sufficiently large m; the inequality uses Lemma 4.5 for the denominator and Lemma 4.7 for the numerator. Sending $m \to \infty$ completes the proof of Proposition 5.12. □

For the result on periodic orbits, let μ_n be the measures defined in (2.12). As in [Wal82, Theorem 9.10], any weak* limit point of the sequence μ_n is invariant with $h(\mu) + \int \varphi \, d\mu \geq \overline{\lim}_{n\to\infty} \frac{1}{n} \log \sum_{x \in \mathrm{Per}_k} e^{S_k\varphi(x)}$. Since every point in $\mathcal{F}$ gives rise to a periodic orbit, this growth rate is at least $P(\mathcal{F},\varphi) = P(\varphi)$; this shows that every weak*-limit point of the sequence μ_n is an equilibrium state for (X,φ), and such limit points exist by compactness. Once uniqueness is established, this will give (iii).

5.4. **Finite multiplicity, liftability, and uniqueness.** In this section we prove (A.5) and (A.6), and use these to deduce uniqueness of the equilibrium state for (X,σ,φ). Note that this is not necessary for Theorem 1.2 since in that setting π is injective and we can put $\nu = (\pi^{-1})_*\mu$; however, for Theorem 1.6 we do not necessarily have injectivity and require the arguments here.

Let μ be any equilibrium state for (X,σ,φ), and assume that I is uniquely decipherable and I^* satisfies $[\mathbf{II}']$; we want to prove that the coding map π has finite multiplicity μ-a.e. To this end, we proceed as in §5.2 and let $\mathbf{a} = (v,1) \in A_I$ be a vertex in the graph giving Σ, where $v \in I$ is arbitrary but fixed, and let $\hat{I} = I \setminus \{v\}$. Consider the collection

$$\mathcal{E}'' := \{\pi(\mathbf{z}_1 \cdots \mathbf{z}_n) \mid \mathbf{z} \in \Sigma, n \in \mathbb{N}, \mathbf{z}_i \neq \mathbf{a} \text{ for all } 1 \leq i \leq n\}$$

of all words in $\mathcal{L}$ that can be lifted to a word in the language of Σ that avoids $\mathbf{a}$. With $\mathcal{E}' = I \cup \mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p\mathcal{F}\mathcal{E}^s)$ as in the last section, observe that $\mathcal{E}'' \subset \mathcal{E}'(\hat{I})^*\mathcal{E}'$, and by Lemma 5.7 it follows that $P(\mathcal{E}',\varphi) < P(\varphi)$. Now consider the sets $B_N' = \{x \in X \mid x_{[-\ell,k-\ell)} \in \mathcal{E}'' \text{ for some } k \geq N \text{ and } \ell \in [0,k]\}$, as in (5.10). Note that the only property of $\mathcal{E}'$ used in Lemma 5.10 was that $P(\mathcal{E}',\varphi) < P(\varphi)$, so Lemma 5.10 applies to the sets B_N'; in particular, there is $N \in \mathbb{N}$ such that $\mu(X \setminus B_N') > 0$. Consider the set $E := \{x \in X \mid f^n(x) \in B_N \text{ for infinitely many positive } n \text{ and infinitely many negative } n\}$. Poincaré recurrence implies that $\mu(E) = 1$, and so (A.5) is a consequence of the following lemma, the idea of which goes back to Bowen [Bow78, p. 12–13] (see also [PP90, p. 229] and [Sar13, Theorem 12.8]).

Lemma 5.13. *For every $x \in E$ we have $\#\pi^{-1}(x) \le N^2$.*

Proof. Suppose there are N^2+1 distinct points $\mathbf{z}^1, \dots, \mathbf{z}^{N^2+1} \in \Sigma$ such that $\pi(\mathbf{z}^i) = x$ for every $1 \le i \le N^2+1$. Then there is $m \in \mathbb{N}$ such that the words $\mathbf{z}^i_{[-m,m]}$ are all distinct. By the definition of E, there are $n_1, n_2 \in \mathbb{Z}$ such that $n_1 \le -m < m \le n_2$ and $f^{n_1}(x), f^{n_2}(x) \in B'_N$. By unique decipherability and the definition of Σ, for every pair of integers $k_1 \in (n_1 - N, n_1]$ and $k_2 \in [n_2, n_2+N)$ there is at most one $\mathbf{z} \in \pi^{-1}(x)$ such that $\mathbf{z}_{k_1} = \mathbf{z}_{k_2} = \mathbf{a}$. In particular, from the collection $\mathbf{z}^1, \dots, \mathbf{z}^{N^2+1}$, there must be some $\mathbf{z}^i \in \pi^{-1}(x)$ with the property that $\mathbf{z}^i_k \ne \mathbf{a}$ for every $k \in (n_1 - N, n_1]$, or $\mathbf{z}^i_k \ne \mathbf{a}$ for every $k \in [n_2, n_2+N)$. But then we either have $\pi(\mathbf{z}^i_{(n_1-N,n_1]}) \in \mathcal{E}''$ or $\pi(\mathbf{z}^i_{[n_2,n_2+N)}) \in \mathcal{E}''$, contradicting the assumption that $f^{n_1}(x), f^{n_2}(x) \in B'_N$. $\square$

Now we prove (A.6) via the following lemma.

Lemma 5.14. *If μ is an ergodic invariant measure on X such that $\mu(\pi(\Sigma)) = 1$ and $\#\pi^{-1}(x) < \infty$ for μ-a.e. $x \in X$, then there is an ergodic invariant measure ν on Σ such that $\pi_*\nu = \mu$, and moreover $h(\nu) = h(\mu)$.*

Proof. This is contained in the proof of [Sar13, Proposition 13.2]. We give the outline here and refer to [Sar13] for details. Define a measure $\tilde{\mu}$ on Σ by

$$\tilde{\mu}(E) := \int_X \left(\frac{1}{\#\pi^{-1}(x)} \sum_{\mathbf{z} \in \pi^{-1}(x)} \mathbf{1}_E(\mathbf{z}) \right) d\mu(x);$$

the idea is to prove that $\tilde{\mu}$ is a well-defined invariant Borel probability measure on Σ, and that almost every ergodic component ν of $\tilde{\mu}$ satisfies the conclusion of the lemma. Although [Sar13, Proposition 13.2] is stated in the context of surface diffeomorphisms, the proof only uses the fact that μ-a.e. point has at least one and at most finitely many preimages. The fact that $h(\nu) = h(\mu)$ is a standard result on finite extensions; for a proof see [NP66, Lemma 1 and Corollary] or [Buz99, Proposition 2.8]. $\square$

To deduce uniqueness, we observe that by Lemma 5.14, every equilibrium state for (X, φ) has the form $\mu = \pi_*\nu$, where ν is an equilibrium state for $(\Sigma, \Phi = \varphi \circ \pi)$; this is since $P_G(\Sigma, \Phi) \ge h(\nu) + \int \Phi \, d\nu = h(\mu) + \int \varphi \, d\mu = P(X, \varphi)$, and the first and last expressions are equal by (5.4). From §5.2, unique decipherability and **[II′]** imply existence of a unique equilibrium state for (Σ, Φ), and we conclude that (X, φ) has a unique equilibrium state, establishing (i). The discussion in §5.3 gives (ii) and (iii).

5.5. **Statistical properties using strong positive recurrence.** Now we deduce statistical properties for the unique equilibrium state. We follow the formulation given by Cyr and Sarig in [CS09] which is most convenient to our present setting. What follows could also be done using the machinery of Young towers developed in [You98, You99].

Consider the one-sided shift (Σ^+, T) with the potential function Φ^+ and the invariant measure m from §5.2. We saw there that (Σ^+, T, Φ^+) is strongly positive recurrent, and by [CS09, Theorem 2.1], this implies the **spectral gap property** [CS09, Definition 1.1]. Then [CS09, Theorem 1.1] implies the following.

- *Exponential decay of correlations.* There is $\theta \in (0,1)$ such that for every $\Psi_1 \in L^\infty(\Sigma^+, m)$ and every bounded $\Psi_2 \in C_\beta(\Sigma^+)$, there is $K = K(\Psi_1, \Psi_2) > 0$ such that for every $n \in \mathbb{N}$ we have

$$\operatorname{Cor}_n^m(\Psi_1, \Psi_2) = \left| \int (\Psi_1 \circ T^n)\Psi_2 \, dm - \int \Psi_1 \, dm \int \Psi_2 \, dm \right| \leq K\theta^n. \tag{5.14}$$

- *Central limit theorem.* If $\Psi \in C_\beta(\Sigma^+)$ is bounded with $\int \Psi \, dm = 0$ and is not cohomologous to a constant, then there is $\sigma_\Psi > 0$ such that for every $\tau \in \mathbb{R}$ we have

$$\lim_{n\to\infty} \mu \left\{ \mathbf{z} \in \Sigma^+ \mid \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \Psi(T^k x) \leq \tau \right\} = \frac{1}{\sigma_\Psi \sqrt{2\pi}} \int_{-\infty}^{\tau} e^{-\frac{t^2}{2\sigma_\Psi^2}} \, dt. \tag{5.15}$$

- *Analyticity of pressure.* If $\Psi \in C_\beta(\Sigma^+)$ is bounded, then $t \mapsto P_G(\Phi^+ + t\Psi)$ is real analytic on a neighbourhood of 0.

We use these results to prove (iv)–(vi) for X and X^+. (Note that the Bernoulli property in (iv) follows from [Sar11].)

For exponential decay of correlations, we first go from Σ^+ to X^+, then obtain the result for X via a standard approximation argument. Recall that $p\colon X \to X^+$ is the map that takes a bi-infinite sequence to its forward infinite half, and let $\bar\mu = p_*\mu$. By commutativity of the diagrams in (5.1), we have $\bar\mu = \pi^+_* m$. Thus for any $\psi_1 \in L^\infty(X^+, \bar\mu)$ and $\psi_2 \in C_\beta(X^+)$, we can put $\Psi_i = \psi_i \circ \pi^+$ and use (5.14) to get

$$\begin{aligned} \operatorname{Cor}_n^{\bar\mu}(\psi_1, \psi_2) &= \left| \int (\psi_1 \circ \sigma^n)\psi_2 \, d\bar\mu - \int \psi_1 \, d\bar\mu \int \psi_2 \, d\bar\mu \right| \\ &= \left| \int (\Psi_1 \circ T^n)\Psi_2 \, dm - \int \Psi_1 \, dm \int \Psi_2 \, dm \right| \leq K\theta^n. \end{aligned} \tag{5.16}$$

This proves exponential decay of correlations for $(X^+, \sigma, \bar\mu)$.

Exponential decay of correlations for the two-sided shift (X, σ, μ) follows from the one-sided result via a standard argument; see [PP90, Proposition 2.4] or [You98, §4]. Roughly speaking, the idea is to approximate $\psi_1, \psi_2 \in C_\beta(X)$ with ψ_1^k, ψ_2^k that depend only on coordinates $-k, \dots, k$ (for example, one can obtain ψ_i^k as a conditional average of ψ_i over $[-k,k]$-cylinders).[22] In [PP90] this is done for functions on X, while in [You98] it is done at the level of the tower; although the notation is different there, the idea is that one takes a function $\psi\colon X \to \mathbb{R}$, considers $\Psi = \psi \circ \pi\colon \Sigma \to \mathbb{R}$, and then approximates Ψ with Ψ_k. Ultimately one is able to reduce to the one-sided

[22] In either case, one must ensure that a single constant K can be chosen to work for all k in (5.14) and (5.16).

case; we omit the details as there is nothing new in our setting. Note that for the two-sided result one must assume that *both* test functions are Hölder.

For the central limit theorem, we observe first that every $\psi \in C_\beta(X^+)$ can also be considered as a function $X \to \mathbb{R}$, and so it suffices to prove the CLT for (X, σ, μ). To this end, consider $\psi \in C_\beta(X)$ with $\int \psi \, d\mu = 0$. Let $\Psi = \psi \circ \pi \in C_\beta(\Sigma)$, and let Ψ^+, u be as in Lemma 5.5. If Ψ^+ is cohomologous to a constant, then Ψ is as well, so there is $f \in L^2(\Sigma, \pi_*^{-1}\mu)$ such that $\Psi = f - f \circ T$. Define $g \in L^2(\Sigma, \mu)$ by $g(x) = f(\pi^{-1}x)$ on $\pi(\Sigma)$, and $g = 0$ elsewhere; then $\psi = g - g \circ \sigma$ on $\pi(\Sigma)$, so ψ is cohomologous to a constant.

Thus if ψ is not cohomologous to a constant, then Ψ is not either, so (5.15) holds for some $\sigma_\Psi > 0$. By (5.2) we have

$$\left| \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \Psi^+(T^k \hat{p}(\mathbf{z})) - \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \Psi(T^k \mathbf{z}) \right| \leq \frac{2\|u\|}{\sqrt{n}}. \tag{5.17}$$

Write $G_n^\mu(\tau) = \mu\{x \in X \mid \frac{1}{\sqrt{n}} \sum_{k=0}^{n-1} \psi(\sigma^k x) \leq \tau\}$, and similarly for $G_n^m(\tau)$ (summing over T-orbits on Σ^+). Then (5.17) gives

$$G_n^m\left(\tau - \frac{2\|u\|}{\sqrt{n}}\right) \leq G_n^\mu(\tau) \leq G_n^m\left(\tau + \frac{2\|u\|}{\sqrt{n}}\right),$$

and it follows that $G_n^\mu(\tau)$ converges to the right-hand side of (5.15) (this uses continuity of that expression).

6. Proof of Theorem 3.2

Now we turn our attention to Theorem 3.2 and assume that we have $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s \subset \mathcal{L}$ satisfying **[I]**, **[II]**, **[III$_\mathrm{a}$]**, **[III$_\mathrm{b}$]**. We must produce $\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s \subset \mathcal{L}$ satisfying **[I$_0$]**, **[II′]**, and **[III*]**. We first dispense with a trivial case where everything reduces to a single periodic orbit. We say that $\mathcal{G} \subset \mathcal{L}$ is **periodic** if there is a periodic sequence $x \in X$ such that every $w \in \mathcal{G}$ appears somewhere in x.

Proposition 6.1. *If (X, φ) is such that there is a periodic $\mathcal{G} \subset \mathcal{L}$ satisfying* **[I]** *and* **[II]***, then there are $\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s \subset \mathcal{L}$ satisfying* **[I$_0$]***,* **[II′]***, and* **[III*]***.*

Proof. Let $x \in X$ be periodic such that every $w \in \mathcal{G}$ appears in x. Let $d \in \mathbb{N}$ be the least period of x and let $\mathcal{F} = \{x_{[1,kd]} \mid k \in \mathbb{N}\}$. Then $\mathcal{F}$ has **[I$_0$]** and **[III*]** (the second assertion uses the fact that d is minimal). Let

$$\mathcal{E}^p = \mathcal{C}^p \mathcal{L}_{\leq d} \cap \mathcal{L} = \{wv \in \mathcal{L} \mid w \in \mathcal{C}^p, |v| \leq d\},$$
$$\mathcal{E}^s = \mathcal{L}_{\leq d} \mathcal{C}^s \cap \mathcal{L} = \{vw \in \mathcal{L} \mid |v| \leq d, w \in \mathcal{C}^s\}.$$

Then given any $u^p \in \mathcal{C}^p, v \in \mathcal{G}, u^s \in \mathcal{C}^s$, we note that there are $i \in [1, d]$ and $j \in (|v| - d, |v|]$ such that $v_{[i,j]} \in \mathcal{F}$, and hence

$$u^p v u^s = (u^p v_{[1,i)}) v_{[i,j]} (v_{(j,|v|]} u^s) \in \mathcal{E}^p \mathcal{F} \mathcal{E}^s.$$

Together with the observation that $P(\mathcal{E}^p, \varphi) = P(\mathcal{C}^p, \varphi)$, $P(\mathcal{E}^s, \varphi) = P(\mathcal{C}^s, \varphi)$, and $I = \mathcal{F} \setminus \mathcal{F}\mathcal{F}$ is finite, this establishes **[II′]** for $\mathcal{E}^p$, $\mathcal{F}$, $\mathcal{E}^s$. $\square$

Recall that we will produce $\mathcal{F}$ satisfying $[\mathbf{I}_0]$ by finding a synchronising triple (r, c, s) for $\mathcal{G}$; that is, $r, s \in \mathcal{G}$ and $c \in \mathcal{L}_{\leq \tau}$ such that for every $r' \in \mathcal{L}r \cap \mathcal{G}$ and $s' = s\mathcal{L} \cap \mathcal{G}$ we have $r'cs' \in \mathcal{G}$; then we will take $\mathcal{F} = \{cw \mid w \in \mathcal{L}r \cap s\mathcal{L} \cap \mathcal{G}\}$. By a careful choice of r, s we will also use $[\mathbf{III}_\mathrm{a}]$ for $\mathcal{G}$ to deduce $[\mathbf{III}^*]$ for $\mathcal{F}$. This is the content of the following result, which implies Proposition 3.7 and is proved in §6.1.

Proposition 6.2. *Every $\mathcal{G} \subset \mathcal{L}$ satisfying $[\mathbf{I}]$ has a synchronising triple (r, c, s), and r, s can be chosen to be arbitrarily long. In addition, we have:*

(a) *If (r, c, s) is any synchronising triple for $\mathcal{G}$, then $\mathcal{F}^{r,c,s}$ satisfies $[\mathbf{I}_0]$.*
(b) *If (r, c, s) is any synchronising triple for $\mathcal{G}$, then a measure μ is Gibbs for φ with respect to $\mathcal{G}$ if and only if it is Gibbs with respect to $\mathcal{F}^{r,c,s}$.*
(c) *If $\mathcal{G}$ is not periodic, then it has a synchronising triple (r, c, s) with r, s arbitrarily long and*

$$[rcs] \cap \sigma^{-k}[rcs] = \emptyset \text{ for every } 1 \leq k \leq \max\{|rc|, |cs|\}. \tag{6.1}$$

(d) *If $\mathcal{G}$ satisfies $[\mathbf{III}_\mathrm{a}]$ for some $L \in \mathbb{N}$, and (r, c, s) is any synchronising triple satisfying (6.1) and $|r|, |s| \geq L$, then $\mathcal{F}^{r,c,s}$ satisfies $[\mathbf{III}^*]$.*

After Proposition 6.2 is proved, the rest of the proof of Theorem 3.2 is devoted to establishing $[\mathbf{II}']$ via the following result, which we prove in §6.2.

Proposition 6.3. *Let X be a shift space on a finite alphabet and $\varphi \in C_\beta(X)$ for some $\beta > 0$. Suppose $\mathcal{G} \subset \mathcal{L}(X)$ satisfies $[\mathbf{I}]$, $[\mathbf{II}]$, $[\mathbf{III}_\mathrm{a}]$, $[\mathbf{III}_\mathrm{b}]$. If (r, c, s) is any synchronising triple for $\mathcal{G}$ satisfying (6.1) and $|r|, |s| \geq L$, then $\mathcal{F} = \mathcal{F}^{r,c,s} = c(s\mathcal{L} \cap \mathcal{L}r \cap \mathcal{G})$ satisfies $[\mathbf{II}']$: for $I = \mathcal{F} \setminus \mathcal{F}\mathcal{F}$ we have $P(I, \varphi) < P(\varphi)$, and there are $\mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L}$ such that $P(\mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p\mathcal{F}\mathcal{E}^s), \varphi) < P(\varphi)$.*

6.1. **Producing a collection of words with free concatenation.** In this section we prove Proposition 6.2. We start by establishing existence of arbitrarily long synchronising triples in §6.1.1. In §6.1.2 we prove parts (a) and (b) (condition $[\mathbf{I}_0]$ and equivalence of the Gibbs properties). In §6.1.3 we prove part (c) by constructing a synchronising triple with no short overlaps, so that (6.1) holds. In §6.1.4 we show that (6.1) gives $[\mathbf{III}^*]$, proving (d).

6.1.1. *Existence of a synchronising triple.* The following lemma mimics the proof from [Ber88] that specification implies synchronised.

Lemma 6.4. *Suppose $\mathcal{G}$ satisfies $[\mathbf{I}]$. Then given $v, w \in \mathcal{G}$, there are $q \in v\mathcal{L} \cap \mathcal{G}$, $p \in \mathcal{L}w \cap \mathcal{G}$, and $c \in \mathcal{L}_{\leq \tau}$ such that $pcq \in \mathcal{G}$, and given any $up, qu' \in \mathcal{G}$, we have $upcqu' \in \mathcal{G}$.*

Proof. Let $C(w, v)$ be the set of connecting words $c \in \mathcal{L}_{\leq \tau}$ such that $wcv \in \mathcal{G}$. This is non-empty by $[\mathbf{I}]$. Let $v^0 = v$ and $w^0 = w$, then define v^n, w^n recursively (see Figure 6.1):

- $v^{n+1} \in \mathcal{G} \cap v^n\mathcal{L}$,
- $w^{n+1} \in \mathcal{G} \cap \mathcal{L}w^n$,
- $C(w^{n+1}, v^{n+1}) \neq C(w^n, v^n)$.

Each $C(w^n, v^n)$ is finite, non-empty, and contained in $C(w^{n-1}, v^{n-1})$. Thus the process terminates for some finite value of n. Let $q = v^n$, $p = w^n$, and pick any $c \in C(p, q)$. Then $pcq \in \mathcal{G}$. Moreover, by the construction of v^n, w^n we see that for any $u, u' \in \mathcal{L}$ with $up, qu' \in \mathcal{G}$, we have $C(up, qu') = C(p, q) \ni c$, hence $upcqu' \in \mathcal{G}$. $\square$

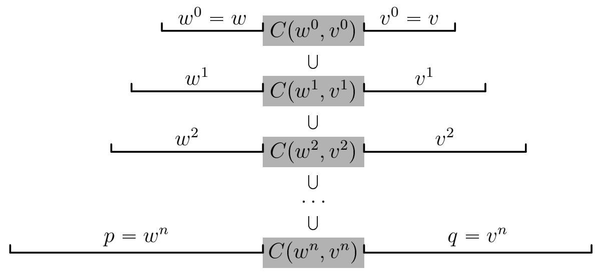


FIGURE 6.1. Producing a synchronising triple.

The triple (p, c, q) produced in Lemma 6.4 is a synchronising triple for $\mathcal{G}$. Note that p, q can be taken arbitrarily long by choosing long words v, w in the lemma.

6.1.2. *Free concatenation and Gibbs properties.* Now we must show that writing $\mathcal{B} = \mathcal{L}p \cap q\mathcal{L} \cap \mathcal{G}$, the collection $c\mathcal{B}$ satisfies $[\mathbf{I}_0]$. This is a consequence of the following lemma.

Lemma 6.5. *If $u^1, \dots, u^n \in \mathcal{B}$, then $u^1 c u^2 c \cdots c u^n \in \mathcal{B}$.*

Proof. The case $n = 1$ is immediate. If the statement holds for n, then for any $u^1, \dots, u^{n+1} \in \mathcal{B}$ we have $u^1 c \cdots c u^n \in \mathcal{B}$. By the definition of $\mathcal{B}$ we have $u^1 c \cdots c u^n = vp \in \mathcal{G}$ for some $v \in \mathcal{L}$, and $u^{n+1} = qw \in \mathcal{G}$ for some $w \in \mathcal{L}$, so

$$u^1 c \cdots c u^n c u^{n+1} = vpcqw.$$

By Lemma 6.4 this is contained in $\mathcal{G}$. Moreover, this word begins with the word q (since u^1 does) and ends with the word p (since u^{n+1} does), so it is an element of $\mathcal{B}$. This establishes the claim for $n+1$, and the result follows by induction. $\square$

Writing $\mathcal{F} = c\mathcal{B}$, we see that for any $v, w \in c\mathcal{B}$ there are $v', w' \in \mathcal{B}$ such that $v = cv'$, $w = cw'$; Lemma 6.5 gives $v'cw' \in \mathcal{B}$, hence $vw = cv'cw' \in c\mathcal{B} = \mathcal{F}$. Thus $\mathcal{F}$ satisfies $[\mathbf{I}_0]$.

For part (b) of Proposition 6.2, we show that a measure μ has the Gibbs property for φ on $\mathcal{F}$ if and only if it has the Gibbs property on $\mathcal{G}$. Note that the upper bound in (2.11) is required to hold for all w, so it suffices to check the lower bound.

Suppose μ is Gibbs on $\mathcal{G}$ with constant Q_1. Then as in Remark 3.5 we have $rw \in \mathcal{G}$ for each $w \in \mathcal{F}^{p,c,q}$, and in particular

$$\begin{aligned}\mu[w] \geq \mu[rw] &\geq Q_1^{-1} e^{-|rw|P(\varphi)+\hat{\varphi}(rw)} \\ &\geq Q_1^{-1} e^{-|w|P(\varphi)+\hat{\varphi}(w)} e^{-|r|(P(\varphi)+\|\varphi\|)},\end{aligned}$$

so μ is Gibbs on $\mathcal{F}^{p,c,q}$. Conversely, if μ is Gibbs on $\mathcal{F}^{p,c,q}$ then given $w \in \mathcal{G}$ there are $u, v \in \mathcal{L}_{\leq \tau}$ such that $cquwvp \in c(q\mathcal{L} \cap \mathcal{L}p \cap \mathcal{G}) = \mathcal{F}^{p,c,q}$, and hence

$$\begin{aligned}\mu[w] \geq \mu[cquwvp] &\geq Q_1^{-1} e^{-|cquwvp|P(\varphi)+\hat{\varphi}(cquwvp)} \\ &\geq Q_1^{-1} e^{-(|cq|+|p|+2\tau)(P(\varphi)+\|\varphi\|)} e^{-|w|P(\varphi)+\hat{\varphi}(w)},\end{aligned}$$

so μ is Gibbs on $\mathcal{G}$. This completes the proof of Proposition 6.2(a)–(b).

6.1.3. *A synchronising triple with no long overlaps.* Now we prove Proposition 6.2(c); assume that $\mathcal{G} \subset \mathcal{L}$ satisfies **[I]** and is not periodic.

Condition (6.1) can be thought of as forbidding 'long overlaps' of rcs with itself (see Figure 6.2 below). To produce a synchronising triple with no long overlaps, we start by letting (p, c, q) be any synchronising triple for $\mathcal{G}$. Note that p, q can be taken arbitrarily long; in particular, they can be taken longer than L from $[\mathbf{III}_{\rm a}]$ (see Lemma 6.4). Then we will take $r = vup$ and $s = qu'w$, where $v, w \in \mathcal{G}$ will be chosen to satisfy certain conditions given below, and $u, u' \in \mathcal{L}_{\leq \tau}$ come from **[I]**. Note that then (r, c, s) is once again a synchronising triple for $\mathcal{G}$.

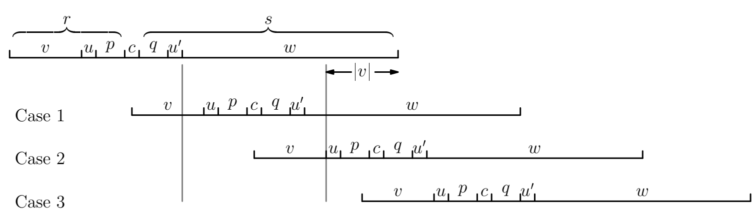


FIGURE 6.2. Cases 1 and 2 will be forbidden by our choice of v, w. Case 3 is permissible.

Before choosing v, w, we suppose r, s have the form just stated, and that $k > 0$ is such that $[rcs] \cap \sigma^{-k}[rcs] \neq \emptyset$. Let x be an element of the intersection, then we have

$$x_{[0,|rcs|)} = x_{[k,k+|rcs|)} = rcs = vupcqu'w. \tag{6.2}$$

We want to choose v, w such that (6.2) forces $k \geq \max(|vupc|, |cqu'w|)$. We will choose v to be (much) longer than p, q, and w to be (much) longer than v. Figure 6.2 illustrates the three possible ranges of k that we must deal with:

(1) $1 \leq k \leq |vupcqu'|$;
(2) $|vupcqu'| < k < |upcqu'w| = |rcs| - |v|$;
(3) $k \geq |upcqu'w| \geq \max(|vupc|, |cqu'w|)$.

Note that if (6.2) holds for some $k < |w|$, then $w_{[1,|w|-k]} = w_{[k+1,|w|]}$, so

$$w_{i+k} = w_i \text{ for every } 1 \le i \le |w| - k. \tag{6.3}$$

Say that a word w satisfying (6.3) is **k-periodic**. Roughly speaking, the idea is that when $k \ll n$, most words in $\mathcal{L}_n$ are not k-periodic, and in particular there is $\alpha > 0$ such that there are arbitrarily long words $w \in \mathcal{G}$ that are not k-periodic for any $k \le \alpha |w|$. Choosing such a w will force $k \ge \alpha |w|$ whenever (6.2) holds, and choosing the lengths of v, w appropriately will guarantee that this rules out Case 1.

Then it will remain to choose $v \in \mathcal{G}$ to rule out Case 2. We will choose v to be short enough that $|vupcqu'| \le \alpha |w|$ (this is necessary for the argument in the previous paragraph for Case 1). Then any k satisfying (6.2) has $i := k - |vupcqu'| > 0$, and it follows from (6.2) that we have either $w_{[i,i+|v|)} = v$ or $i > |w| - |v|$. We will choose v to be a word that does not appear as a subword of w, which will force the inequality $i > |w| - |v|$, so that $k = i + |v| + |upcqu'| \ge |upcqu'w|$, which will suffice, since $|cqu'w| \ge |vupc|$ by our choice of v.

With the above plan in mind, we now carry out the details. We need to guarantee that $\mathcal{G}_n$ contains enough words that we can throw out the 'bad' ones and still have something left. This requires the non-periodicity condition.

Lemma 6.6. *Suppose $\mathcal{G}$ has* **[I]** *and is not periodic, and let c be the connecting word from some synchronising triple for $\mathcal{G}$. Then there is $\ell \in \mathbb{N}$ such that $\#\mathcal{G}_{\ell m - |c|} \ge 2^m$ for all $m \in \mathbb{N}$.*

Remark 6.7. A very similar result is proved in [CT12, Proposition 2.4] (see §6.3 there). Our hypotheses here are weaker (the gluing time in specification is allowed to vary) and the results in the previous section allow us to give a simpler proof. Note that the conclusion is strictly stronger than the inequality $h(\mathcal{G}) > 0$.

Proof of Lemma 6.6. First note that if $\mathcal{G}$ is not periodic, then neither is $c\mathcal{B}$. Indeed, if $x \in A^{\mathbb{N}}$ was periodic and contained every word in $c\mathcal{B}$ as a subword, then it would contain every word in $\mathcal{G}$ as well, by Remark 3.5.

Now we claim that there are $v, w \in c\mathcal{B}$ such that $|w| \ge |v|$ and $w \notin v\mathcal{L}$. Suppose this was false; we will show that in this case $c\mathcal{B}$ must be periodic. Indeed, define $x \in A^{\mathbb{N}}$ by $x_i = w_i$ for some $w \in c\mathcal{B}$ with $|w| \ge i$. By the assumption this is well-defined since $v_i = w_i$ whenever $v, w \in c\mathcal{B}$ and $i \le \min(|v|, |w|)$. Moreover, given $v \in c\mathcal{B}$ we have $vv \cdots v \in c\mathcal{B}$ for arbitrarily long concatenations of v with itself, so $x_{i+k|v|} = v_i = x_i$ for any $k \ge 0$ and $1 \le i \le |v|$. It follows that x is periodic, and any $w \in c\mathcal{B}$ appears as a prefix of $vv \cdots v$, hence as a prefix of x, by the assumption.

The previous paragraph shows that non-periodicity of $c\mathcal{B}$ implies existence of $v, w \in c\mathcal{B}$ such that $v_i \neq w_i$ for some $i \le \min(|v|, |w|)$. Let $\ell = |v| \cdot |w|$; let $u^1 = vv \cdots v$ and $u^2 = ww \cdots w$, where we concatenate $|w|$ copies of v and $|v|$ copies of w so that $|u^1| = |u^2| = \ell$. By construction of v, w we have

$u^1 \neq u^2$. Now for every $m \in \mathbb{N}$ and $y \in \{1,2\}^m$, that is, every finite sequence of 1s and 2s, we have $u^{y_1} \cdots u^{y_m} \in (c\mathcal{B})_{m\ell}$. Moreover, different choices of y yield different words in $(c\mathcal{B})_{m\ell}$, so that $\#(c\mathcal{B})_{m\ell} \geq 2^m$. Since $\mathcal{B} \subset \mathcal{G}$ we get $\#\mathcal{G}_{m\ell-|c|} \geq 2^m$. $\square$

Now we choose the lengths of the words v, w described before Lemma 6.6. Recall that A is the alphabet of X, and let ℓ be as in Lemma 6.6. Choose $\alpha > 0$ such that $\alpha\ell \log(\#A) < \log 2$. Note that we can choose arbitrarily large $m, n \in \mathbb{N}$ such that

$$\frac{\ell}{\alpha} m + \frac{2\tau + |pcq|}{\alpha} \leq \ell n - |c| < 2^m. \tag{6.4}$$

We claim that for sufficiently large m, n satisfying (6.4), there exist $v \in \mathcal{G}_{\ell m-|c|}$ and $w \in \mathcal{G}_{\ell n-|c|}$ such that v is not a subword of w, and w is not k-periodic for any $1 \leq k \leq \alpha\,|w|$.

To this end, we consider the collection

$$\mathcal{P}^\alpha = \{w \in \mathcal{L} \mid w \text{ is } k\text{-periodic for some } 1 \leq k \leq \alpha\,|w|\}.$$

If w is k-periodic then it is determined by its first k entries, so we can estimate the cardinality of $\mathcal{P}^\alpha_N$ by

$$\#\mathcal{P}^\alpha_N \leq \sum_{k=1}^{\lfloor \alpha N \rfloor} (\#A)^k \leq (\#A)^{\alpha N} \sum_{j=0}^{\infty} (\#A)^{-j} = e^{\alpha \log(\#A) N} \left(\frac{\#A}{\#A - 1} \right). \tag{6.5}$$

Write $\gamma := (\log 2)/\ell - \alpha \log \#A$ and note that $\gamma > 0$ by the choice of α. When $N = \ell n - |c|$ for some $n \in \mathbb{N}$, Lemma 6.6 gives $\#\mathcal{G}_N \geq 2^n \geq 2^{N/\ell}$, and so

$$\frac{\#\mathcal{G}_N}{\#\mathcal{P}^\alpha_N} \geq \left(\frac{\#A - 1}{\#A} \right) e^{(\log 2)\frac{N}{\ell} - \alpha \log(\#A) N} \geq \left(\frac{\#A - 1}{\#A} \right) e^{\gamma N}. \tag{6.6}$$

For n sufficiently large this gives $\#\mathcal{G}_N > \#\mathcal{P}^\alpha_N$, so there is $w \in \mathcal{G}_N = \mathcal{G}_{\ell n-|c|}$ that is not k-periodic for any $1 \leq k \leq \alpha\,|w|$. To put it another way: for every sufficiently large n there is $w \in \mathcal{G}_{\ell n-|c|}$ such that

$$\text{for every } 1 \leq k \leq \alpha\,|w| \text{ there is } 1 \leq j \leq |w| - k \text{ with } w_{k+j} \neq w_k. \tag{6.7}$$

Now let m, n be such that (6.4) is satisfied and (6.7) holds for some $w \in \mathcal{G}_{\ell n-|c|}$. Note that w contains at most $|w|$ subwords of length $\ell m - |c|$, while $\#\mathcal{G}_{\ell m-|c|} \geq 2^m > |w|$ by (6.4). Thus there is $v \in \mathcal{G}_{\ell m-|c|}$ such that $w_{[i,i+|v|)} \neq v$ for any $1 \leq i \leq |w| - |v|$; that is, v is not a subword of w.

By **[I]** there are $u, u' \in \mathcal{L}_{\leq\tau}$ such that $vup, qu'w \in \mathcal{G}$. By the first inequality in (6.4) we have

$$|vupcqu'| \leq \ell m + 2\tau + |pcq| \leq \alpha\,|w|\,.$$

It follows that w is not k-periodic for any $1 \leq k \leq |vupcqu'|$. Now as in the discussion prior to Lemma 6.6, we see that any k such that (6.2) holds must fall into one of the three classes described there. The first case described there cannot occur because of the aperiodicity of w. The second case cannot

occur because v is not a subword of w. Thus only the third case can occur, which proves (6.1).

6.1.4. *Absence of long overlaps implies* [**III***]. Now we prove part (d) of Proposition 6.2. Let (r,c,s) be a synchronising triple for $\mathcal{G}$ satisfying (6.1). We show that

$$\mathcal{F} := \mathcal{F}^{r,c,s} = c\mathcal{B}^{r,s} = c(s\mathcal{L} \cap \mathcal{L}r \cap \mathcal{G})$$

satisfies [**III***] if $\mathcal{G}$ satisfies [**III**$_\text{a}$]. Note that $\mathcal{F}$ satisfies [**I**$_0$] by part (a).

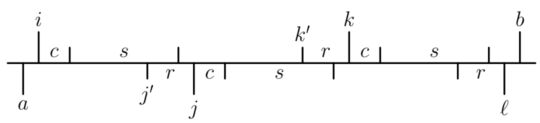


FIGURE 6.3. Establishing [**III***].

Suppose $x \in X$ and $i \leq j < k \leq \ell \in \mathbb{Z}$ are such that $x_{[i,k)}, x_{[j,\ell)} \in \mathcal{F}$ and there are $a < j$ and $b > k$ such that $x_{[a,j)}, x_{[k,b)} \in \mathcal{F}$ (see Figure 6.3). We must show that $x_{[j,k)} \in \mathcal{F}$. Let $j' = j - |r|$ and $k' = k - |r|$; then we have

$$x_{[j',j'+|rcs|)} = x_{[k',k'+|rcs|)} = rcs,$$

so (6.1) gives $k - j = k' - j' \geq \max(|rc|, |cs|)$. Thus $x_{[j,j+|c|)} = c$ and $x_{[j+|c|,k)} \in s\mathcal{L} \cap \mathcal{L}r$.

It remains only to show that $x_{[j+|c|,k)} \in \mathcal{G}$. For this we observe that $x_{[i,k)} \in \mathcal{F}$ implies $x_{[i+|c|,k)} \in \mathcal{G}$, and $x_{[j,\ell)} \in \mathcal{F}$ implies $x_{[j+|c|,\ell)} \in \mathcal{G}$. Note that $i + |c| \leq j + |c| < k \leq \ell$, and that $k - j \geq \max(|rc|, |cs|)$ implies $k - (j + |c|) \geq \max(|r|, |s|) \geq L$, so by [**III**$_\text{a}$] we have $x_{[j+|c|,k)} \in \mathcal{G}$. It follows that $x_{[j,k)} \in \mathcal{F}$, which establishes [**III***] for $\mathcal{F}$. This completes the proof of Proposition 6.2.

6.2. **Construction of $\mathcal{E}^p$ and $\mathcal{E}^s$.** In this section we prove Proposition 6.3. We first give (in §6.2.1) an outline of the proof in the case when [**I**] is satisfied with $\mathcal{G} = \mathcal{L}$ (that is, (X,σ) satisfies the classical specification property, and [**II**], [**III**] are automatic with $\mathcal{C}^p = \mathcal{C}^s = \emptyset$); then in §§6.2.2–6.2.6 we deal with the more general case when there is $\mathcal{G} \subset \mathcal{L}$ satisfying [**I**], [**II**], [**III**$_\text{a}$], [**III**$_\text{b}$]. The argument in §6.2.1 is only sketched because it follows from the stronger result in §6.2.2, but we describe it first for the sake of clarifying the presentation and motivating the extra steps that must be taken in the general case.

6.2.1. *Uniform specification.* Let (r,c,s) be a synchronising triple for $\mathcal{L}$ satisfying (6.1), and let $\mathcal{F} = \mathcal{F}^{r,c,s} = c\mathcal{B}^{r,s} = c(s\mathcal{L} \cap \mathcal{L}r)$, so that $\mathcal{F}$ satisfies [**I**$_0$] and [**III***]. Given $w \in \mathcal{L}$, let

(6.8) $$S(w) := \{i \in [|r|, |w| - |cs|] \cap \mathbb{N} \mid w_{(i-|r|,i+|cs|]} = rcs\}$$

be the collection of **'synchronising times'** at which the synchronising triple appears. In particular, $S(w)$ has the property that $w_{(i,j]} \in \mathcal{F}$ whenever $i, j \in S(w)$ (using **[III*]**, or more directly, (6.1)). Consider the set of words

$$\mathcal{E} := \{w \in \mathcal{L} \mid S(w) = \emptyset\} = \mathcal{L} \setminus \mathcal{L}rcs\mathcal{L}. \tag{6.9}$$

Then given $w \in \mathcal{L} \setminus \mathcal{E}$ we can take $i = \min S(w)$ and $j = \max S(w)$ to get

$$w_{[1,i]} \in \mathcal{E}, \quad w_{(i,j]} \in \mathcal{F}, \quad w_{(j,|w|]} \in \mathcal{E}.$$

Thus taking $\mathcal{E}^p = \mathcal{E}^s = \mathcal{E}$, we have $\mathcal{L} \setminus (\mathcal{E}^p\mathcal{F}\mathcal{E}^s) \subset \mathcal{E}$. Moreover, if $w \in \mathcal{F}$ and $S(w) \neq \emptyset$, then for any $i \in S(w)$ we have $w_{[1,i]}, w_{(i,|w|]} \in \mathcal{F}$, so $w \in \mathcal{F}\mathcal{F}$ and hence $w \notin I := \mathcal{F} \setminus \mathcal{F}\mathcal{F}$. Thus we have shown that

$$I \cup \mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p\mathcal{F}\mathcal{E}^s) \subset \mathcal{E},$$

and so to prove **[II′]** it suffices to show that $P(\mathcal{E}, \varphi) < P(\varphi)$. To this end, fix $T \in \mathbb{N}$ large, and given $0 \leq kT \leq n$, let

$$\mathcal{A}_n^k = \{w \in \mathcal{L}_n \mid I_j \cap S(w) = \emptyset \text{ for every } 0 \leq j < k\}$$

be the set of words that **avoid** rcs for each of the first k intervals of length T. (We ignore instances of rcs that cross the boundary between two of these intervals.) In particular, $\mathcal{A}_n^0 = \mathcal{L}_n$ and $\mathcal{A}_n^k \supset \mathcal{E}$ for all k, so we can control $\Lambda_n(\mathcal{E}, \varphi)$ by estimating $\Lambda_n(\mathcal{A}_n^k, \varphi)$ iteratively (in k). To this end, let $\mathcal{Z}^k = \mathcal{A}_{kT}^k$, so that writing $\mathcal{H}(v, 1) = \mathcal{L} \cap v\mathcal{L}$ as in (4.21), we have

$$\mathcal{A}^k = \bigsqcup_{v \in \mathcal{Z}^k} \mathcal{H}(v, 1). \tag{6.10}$$

For each $v \in \mathcal{Z}^k$, specification gives $q \in \mathcal{L}_{\leq \tau}$ such that $vqrcs \in \mathcal{L}$, and as long as $T \geq \tau + |rcs|$, we get $\mathcal{H}(vqrcs, 1) \subset \mathcal{H}(v, 1) \setminus \mathcal{A}^{k+1}$. The Gibbs bounds from Proposition 4.6 can be used (together with some standard distortion estimates) to get $\gamma > 0$ such that

$$\Lambda_n(\mathcal{H}(v, 1) \setminus \mathcal{A}^{k+1}, \varphi) \geq \gamma \Lambda_n(\mathcal{H}(v, 1), \varphi)$$

for every $v \in \mathcal{Z}^k$. Rewriting as $\Lambda_n(\mathcal{H}(v, 1) \cap \mathcal{A}^{k+1}, \varphi) \leq (1-\gamma)\Lambda_n(\mathcal{H}(v, 1), \varphi)$ and summing over all $v \in \mathcal{Z}^k$, (6.10) gives

$$\Lambda_n(\mathcal{A}^{k+1}, \varphi) \leq (1 - \gamma)\Lambda_n(\mathcal{A}^k, \varphi).$$

In particular, for $k_n = \lfloor \frac{n}{T} \rfloor$, we get

$$\Lambda_n(\mathcal{E}, \varphi) \leq \Lambda_n(\mathcal{A}^{k_n}, \varphi) \leq (1 - \gamma)^{k_n} \Lambda_n(\mathcal{A}^0, \varphi) \leq (1 - \gamma)^{\lfloor \frac{n}{T} \rfloor} Q_2 e^{nP(\varphi)},$$

and deduce that $P(\mathcal{E}, \varphi) \leq P(\varphi) + \frac{1}{T} \log(1 - \gamma) < P(\varphi)$, as desired.

6.2.2. *Non-uniform specification.* In the non-uniform setting we still follow the basic plan from §6.2.1, but a number of complications arise. Suppose $\mathcal{G}$ satisfies **[I]**, **[II]**, **[III$_\mathrm{a}$]**, **[III$_\mathrm{b}$]** and let r, c, s be a synchronising triple satisfying (6.1) and such that $|r|, |s| \geq L$. Let $\mathcal{F} = \mathcal{F}^{r,c,s}$ and $I = \mathcal{F} \setminus \mathcal{F}\mathcal{F}$; we must prove that $P(I, \varphi) < P(\varphi)$ and produce $\mathcal{E}^p, \mathcal{E}^s \subset \mathcal{L}$ such that

$$P(\mathcal{E}^p \cup \mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{E}^p \mathcal{F} \mathcal{E}^s), \varphi) < P(\varphi). \tag{6.11}$$

To this end, given $w \in \mathcal{G}$, consider the set of times where the synchronising triple appears in a 'good' position relative to the start of the word:

$$G_\mathrm{s}^-(w) := \{i \in [1, |w|] \mid w_{[1,i]} \in \mathcal{G} \text{ and } w_{(i-|r|, i+|cs|]} = rcs\}. \tag{6.12}$$

Let $S(w) = \{i \in G_\mathrm{s}^-(w) \mid w_{(i+|c|,|w|]} \in \mathcal{G}\}$ be the set of such times that are also 'good' relative to the end of the word. Then consider the collection

$$\mathcal{E} := \{w \in \mathcal{G} \mid S(w) = \emptyset\}.$$

Let $\mathcal{E}^p := \mathcal{C}^p\mathcal{E}$ and $\mathcal{E}^s := \mathcal{E}\mathcal{C}^s$, so $P(\mathcal{E}^p \cup \mathcal{E}^s, \varphi) \leq \max\{P(\mathcal{E}, \varphi), P(\mathcal{C}^p, \varphi), P(\mathcal{C}^s, \varphi)\}$ by (4.13). Moreover, $I = \mathcal{F} \setminus \mathcal{F}\mathcal{F} \subset \mathcal{E}$, so it remains to describe $\mathcal{L} \setminus \mathcal{E}^p\mathcal{F}\mathcal{E}^s$, and then to estimate $P(\mathcal{E}, \varphi)$.

By **[III$_\mathrm{a}$]** and (6.1), given $w \in \mathcal{G}$ and $i, j \in S(w)$, we have $w_{(i,j]} \in \mathcal{F}$. Given $v \in \mathcal{C}^p\mathcal{G}\mathcal{C}^s$ we write $v = u^p w u^s$ for $u^p \in \mathcal{C}^p$, $w \in \mathcal{G}$, and $u^s \in \mathcal{C}^s$. Then we have the following dichotomy: either

- $\#S(w) \leq 1$, so $w \in \mathcal{E}\mathcal{E}$ and $v \in \mathcal{C}^p\mathcal{E}\mathcal{E}\mathcal{C}^s \subset \mathcal{E}^p\mathcal{E}^s$; or
- $\#S(w) > 1$, in which case we take $i = \min S(w)$ and $j = \max S(w)$ to obtain $w_{[1,i]} \in \mathcal{E}$, $w_{(i,j]} \in \mathcal{F}$, and $w_{(j,|w|]} \in \mathcal{E}$ as in §6.2.1. In particular, this gives $v \in \mathcal{C}^p\mathcal{E}\mathcal{F}\mathcal{E}\mathcal{C}^s \subset \mathcal{E}^p\mathcal{F}\mathcal{E}^s$.

We conclude that $\mathcal{L} \setminus \mathcal{E}^p\mathcal{F}\mathcal{E}^s \subset \mathcal{E}^p\mathcal{E}^s \cup (\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s)$. By **[II]** and the above estimates, in order to prove **[II′]** it suffices to show that $P(\mathcal{E}, \varphi) < P(\varphi)$.

Let $\mathcal{G}^+ := \{w_{[1,i]} \mid w \in \mathcal{G}, 1 \leq i \leq |w|\}$, and similarly $\mathcal{G}^- := \{w_{[i,|w|]} \mid 1 \leq i \leq |w|\}$. The following lemma (proved in §6.2.3) says that with very few exceptions, words in $\mathcal{G}^+$ admit a decomposition with no prefix, and words in $\mathcal{G}^-$ admit a decomposition with no suffix.

Lemma 6.8. *With $\mathcal{C}^p, \mathcal{G}, \mathcal{C}^s$ satisfying* **[I]**, **[II]**, *and* **[III$_\mathrm{a}$]**, *we have $P(\mathcal{G}^+ \setminus (\mathcal{G}\mathcal{C}^s), \varphi) < P(\varphi)$, and similarly $P(\mathcal{G}^- \setminus (\mathcal{C}^p\mathcal{G}), \varphi) < P(\varphi)$. In particular, there are $\xi > 0$ and $Q_7 > 0$ such that for every $n \in \mathbb{N}$ we have*

$$\begin{aligned} \Lambda_n(\mathcal{G}^+ \setminus (\mathcal{G}\mathcal{C}^s), \varphi) &\leq Q_7 e^{n(P(\varphi)-\xi)}, \\ \Lambda_n(\mathcal{G}^- \setminus (\mathcal{C}^p\mathcal{G}), \varphi) &\leq Q_7 e^{n(P(\varphi)-\xi)}. \end{aligned} \tag{6.13}$$

Given $w \in \mathcal{G}^+$, define $G_\mathrm{s}^-(w)$ as in (6.12). For $M \in \mathbb{N}$ and $w \in \mathcal{G}^+$, let

$$G_\mathrm{s}^M(w) := \{i \in [1, |w|] \mid i - |c| \in G_\mathrm{s}^-(w) \text{ and } w_{(i,i+M]} \in \mathcal{G}\},$$

as illustrated in Figure 6.4.

The indices $i \in G_\mathrm{s}^M(w)$ are 'candidates' for good occurrences of the synchronising triple (r, c, s). Note that here i represents the position where s starts, rather than the position where r ends; this differs from our convention with $G_\mathrm{s}^-(w)$ but will be more convenient in what follows. To guarantee

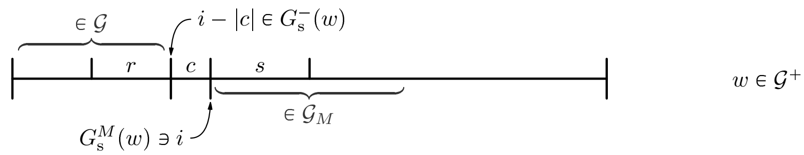


FIGURE 6.4. Candidates for good occurrences of (r, c, s).

that an index $i \in G_s^M(w)$ represents a genuinely good occurrence (that is, $i - |c| \in S(w)$), we will eventually need the added property that there is no long obstruction beginning at i; that is, $w_{(i,i']} \notin \mathcal{C}^p$ for any $i' \geq i + M - L$. By Lemma 6.8 this will (typically) guarantee existence of $j \in (i, i + M - L]$ such that $w_{(j,|w|]} \in \mathcal{G}$; this in turn will allow us to apply [$\mathbf{III}_\mathrm{b}$] and deduce that $w_{(i,|w|]} \in \mathcal{G}$ and hence $i - |c| \in S(w)$.

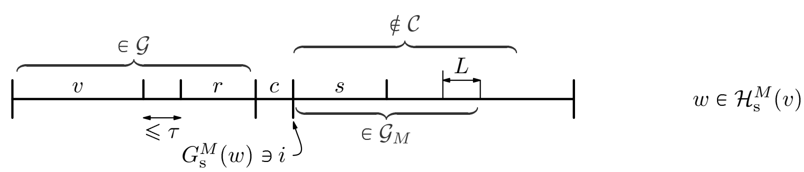


FIGURE 6.5. Obtaining $i \in S(w)$ from $i \in G_s^M(w)$.

To show that nearly every word w has indices in $G_s^M(w)$, we will work left to right, estimating the probability that the synchronising word appears soon in a good position, conditioned on the symbols we have seen so far. Write $\mathcal{C} := \mathcal{C}^p \cup \mathcal{C}^s \cup (\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s)$. For each $v \in \mathcal{G}$, let $\mathcal{H}^+(v) := \mathcal{G}^+ \cap v\mathcal{L}$ and consider the collection

$$\begin{aligned}\mathcal{H}_s^M(v) := \{w \in \mathcal{H}^+(v) \mid \text{there is } i \in [|v| + |rc|\,, |v| + |rc| + \tau] \cap G_s^M(w) \\ \text{such that } w_{(i,i']} \notin \mathcal{C} \text{ for any } i' \in [i + M - L, |w|]\},\end{aligned}$$

illustrated in Figure 6.5. The following is proved in §6.2.4 and is the source of the exponential decay that we eventually obtain.

Lemma 6.9. *There is $\gamma > 0$ such that there are arbitrarily large values of $M \in \mathbb{N}$ such that the following holds for every $v \in \mathcal{G}$ and $n \geq |v| + \tau + |rc| + M$:*

$$\Lambda_n(\mathcal{H}_s^M(v), \varphi) \geq \gamma \Lambda_n(\mathcal{H}^+(v), \varphi). \tag{6.14}$$

In order to apply Lemma 6.9, we need to show that for $w \in \mathcal{G}^+$, we have $w_{[1,i]} \in \mathcal{G}$ 'often enough'. Thanks to Lemma 6.8, we can do this by controlling how many words can have long segments covered by a small number of words in $\mathcal{C}$. To this end, given $\delta, \beta > 0$, consider the collection

$$\begin{aligned}\hat{\mathcal{C}} = \hat{\mathcal{C}}^{\beta,\delta} := \{w \in \mathcal{L} \mid \text{there are } \{(i_a, i'_a] \subset [1, |w|]\}_{a=1}^B \text{ with } B \leq 2\beta\,|w| \\ \text{such that } w_{(i_a,i'_a]} \in \mathcal{C} \text{ and } \#\mathbb{N} \cap \textstyle\bigcup_{a=1}^B (i_a, i'_a] \geq 2\delta\,|w|\}\end{aligned}$$

of words for which at least 2δ of the length of the word can be covered by a small ($< 2\beta |w|$) number of subwords lying in $\mathcal{C}$. When $\beta \ll \delta$, the following estimate (proved in §6.2.5) shows that $\hat{\mathcal{C}}$ has small pressure, and gives a concrete estimate on the partition sum.

Lemma 6.10. *For every $\delta > 0$ there is $\beta > 0$ such that $P(\hat{\mathcal{C}}^{\beta,\delta}, \varphi) < P(\varphi)$. In particular, there are $Q_8 > 0$ and $\theta < 1$ such that*

$$\Lambda_n(\hat{\mathcal{C}}^{\beta,\delta}, \varphi) \leq Q_8 \theta^n e^{nP(\varphi)} \tag{6.15}$$

for every $n \in \mathbb{N}$. Moreover, because $\hat{\mathcal{C}}^{\beta',\delta} \subset \hat{\mathcal{C}}^{\beta,\delta}$ for every $0 < \beta' < \beta$, (6.15) remains true if β is replaced with any smaller positive number.

Given $M \in \mathbb{N}$, put $T = 2(M + |rc| + \tau)$. As in §6.2.1, we will estimate the proportion of words that 'reset' with the synchronising triple in a good position by time kT for each $k \in \mathbb{N}$. Let $\delta = \frac{1}{8}$, and let $\beta > 0$ be such that Lemma 6.10 holds. Choose $M \geq \max(\beta^{-1}, 2(|rc| + \tau))$ such that Lemma 6.9 holds, and such that in addition, $\zeta := (\max(\theta, e^{-\xi}))^T$ satisfies[23]

$$Q_2(Q_7 + Q_8)\zeta(1-\theta)^{-1}(1-\zeta)^{-1} < \gamma. \tag{6.16}$$

Observe our choice of M guarantees that $\frac{1}{2T} < \beta$ and that $T \leq 3M$; we will use this in Lemma 6.11 below. For every $k \in \mathbb{N}$ and $m \geq kT$, Lemmas 6.8 and 6.10 give

$$\begin{aligned} \Lambda_m(\mathcal{G}^+ \setminus (\mathcal{G}\mathcal{C}), \varphi) &\leq Q_7 \zeta^k e^{mP(\varphi)}, \\ \Lambda_m(\hat{\mathcal{C}}^{\beta,\delta}, \varphi) &\leq Q_8 \zeta^k e^{mP(\varphi)}. \end{aligned} \tag{6.17}$$

Consider the intervals $I_j := (jT, (j + \frac{1}{2})T]$ for $j \in \mathbb{N}$, and write

$$\begin{aligned} \mathcal{A}^k := \{w \in \mathcal{G}^+ \mid \text{for all } 0 \leq j < k \text{ and } i \in I_j \cap G_{\mathrm{s}}^M(w), \\ \text{there is } i' \in [i + M - L, |w|] \text{ such that } w_{(i,i']} \in \mathcal{C}\} \end{aligned} \tag{6.18}$$

for the set of words which **avoid** genuinely good occurrences of (r, c, s) in the first k intervals I_j; in particular, any candidate good occurrence within these intervals is 'ruined' sometime before the end of w by a long obstruction from $\mathcal{C}$. Note that $\mathcal{A}^{k+1} \subset \mathcal{A}^k$, and that if $v \in \mathcal{A}^k$ and $|v| \geq kT$, then $\mathcal{H}^+(v) \subset \mathcal{A}^k$. On the other hand, there may be $w \in \mathcal{A}^k$ with $w_{[1,kT]} \notin \mathcal{A}^k$.

Our goal is to relate $\Lambda_n(\mathcal{A}^{k+1}, \varphi)$ and $\Lambda_n(\mathcal{A}^k, \varphi)$. We do this by decomposing (most of) $\mathcal{A}^k$ into collections $\mathcal{H}^+(v)$ where $v \in \mathcal{G}$, and then applying Lemma 6.9. Given $k \in \mathbb{N}$, write $m_k := (k + \frac{1}{2})T - (\tau + |rc|)$ and let

$$\mathcal{Z}^k := \{v \in \mathcal{G} \mid |v| \in [kT, m_k], v_{[1,kT]} \in \mathcal{A}^k, v_{[1,i]} \notin \mathcal{G} \text{ for any } kT \leq i < |v|\}.$$

We will use the following observations.

(1) The collections $\{\mathcal{H}^+(v) \mid v \in \mathcal{Z}^k\}$ are disjoint (no word in $\mathcal{Z}^k$ is a prefix of any other).

[23] Recall that γ is independent of the choice of M.

(2) Given $v \in \mathcal{Z}^k$ and $n \geq (k+1)T$, we have $\mathcal{H}_{\mathrm{s}}^M(v)_n \subset \mathcal{H}_n^+(v) \setminus \mathcal{A}^{k+1}$ (this motivates our choice of m_k).

The following lemma, proved in §6.2.6, relates $\mathcal{A}^k$ to $\bigsqcup_{v\in\mathcal{Z}^k} \mathcal{H}^+(v)$ up to a small term whose partition sum is well controlled by Lemmas 6.8 and 6.10.

Lemma 6.11. *For every $k \geq 0$ and $n \geq m_k$, we have*

$$\bigsqcup_{v\in\mathcal{Z}^k} \mathcal{H}_n^+(v) \subset \mathcal{A}_n^k \subset \Big(\bigsqcup_{v\in\mathcal{Z}^k} \mathcal{H}_n^+(v)\Big) \cup \mathcal{Y}_n^k, \tag{6.19}$$

where

$$\mathcal{Y}_n^k := (\mathcal{G}^+ \setminus (\mathcal{G}\mathcal{C}^s))_{m_k}\mathcal{L}_{n-m_k} \cup \Big(\bigcup_{j=0}^{k-1}\bigcup_{\ell=kT}^{n} \mathcal{A}_{jT}^j \hat{\mathcal{C}}_{\ell-jT}^{\beta,\delta}\mathcal{L}_{n-\ell}\Big),$$

satisfies the bound (with θ, ζ as in (6.15)*–*(6.17)*)*

$$\Lambda_n(\mathcal{Y}^k,\varphi) \leq Q_2 \cdot \frac{Q_7+Q_8}{1-\theta}\sum_{j=0}^{k-1}\Lambda_{jT}(\mathcal{A}^j,\varphi)\zeta^{k-j}e^{(n-jT)P(\varphi)}. \tag{6.20}$$

Now we estimate $\Lambda_n(\mathcal{A}^k,\varphi)$. Given $v \in \mathcal{Z}^k$ and $n \geq (k+1)T$, we recall from above that $\mathcal{H}_{\mathrm{s}}^M(v)_n \subset \mathcal{H}_n^+(v) \setminus \mathcal{A}_n^{k+1}$, and so Lemma 6.9 gives

$$\Lambda_n(\mathcal{H}^+(v)\cap\mathcal{A}^{k+1},\varphi) \leq \Lambda_n(\mathcal{H}^+(v)\setminus\mathcal{H}_{\mathrm{s}}^M(v),\varphi) \leq (1-\gamma)\Lambda_n(\mathcal{H}^+(v),\varphi).$$

Summing over $v \in \mathcal{Z}^k$ and using $\mathcal{A}^{k+1} \subset \mathcal{A}^k$ together with the second inclusion in (6.19), we get the estimate

$$\Lambda_n(\mathcal{A}^{k+1},\varphi) \leq \Big(\sum_{v\in\mathcal{Z}^k}\Lambda_n(\mathcal{H}^+(v)\cap\mathcal{A}^{k+1},\varphi)\Big) + \Lambda_n(\mathcal{Y}^k,\varphi). \tag{6.21}$$

Using the first inclusion in (6.19), the sum in brackets is bounded above by $\sum_{v\in\mathcal{Z}^k}(1-\gamma)\Lambda_n(\mathcal{H}^+(v),\varphi) \leq (1-\gamma)\Lambda_n(\mathcal{A}^k,\varphi)$, and so (6.21) gives

$$\Lambda_n(\mathcal{A}^{k+1},\varphi) \leq (1-\gamma)\Lambda_n(\mathcal{A}^k,\varphi) + \Lambda_n(\mathcal{Y}^k,\varphi). \tag{6.22}$$

Let $a_k := \sup_{n\geq kT}\Lambda_n(\mathcal{A}^k,\varphi)e^{-nP(\varphi)}$ so that $\Lambda_n(\mathcal{A}^k,\varphi) \leq a_k e^{nP(\varphi)}$ for all $n \geq kT$. A priori we allow $a_k = \infty$, but note that $a_0 \leq Q_2$ by Lemma 4.5; we will prove inductively that $a_k \leq Q_2\eta^k$ for some $\eta < 1$. Write (6.20) as

$$\Lambda_n(\mathcal{Y}^k,\varphi) \leq Q_2\frac{Q_7+Q_8}{1-\theta}\sum_{j=0}^{k-1}a_j\zeta^{k-j}e^{nP(\varphi)},$$

and multiply both sides of (6.22) by $e^{-nP(\varphi)}$ to get

$$a_{k+1} \leq (1-\gamma)a_k + Q_2\frac{Q_7+Q_8}{1-\theta}\sum_{j=0}^{k-1}a_j\zeta^{k-j}. \tag{6.23}$$

Rewriting (6.16) as $(1-\gamma)+Q_2(Q_7+Q_8)\zeta(1-\theta)^{-1}(1-\zeta)^{-1} < 1$, we can choose $\eta < 1$ such that

$$(1-\gamma)\eta^{-1} + Q_2\frac{Q_7+Q_8}{1-\theta}\cdot\frac{\zeta}{\eta(\eta-\zeta)} < 1. \tag{6.24}$$

Let $b_k = a_k \eta^{-k}$ and observe that $b_0 = a_0 \le Q_2$. Suppose that k is such that $b_j \le Q_2$ for every $0 \le j \le k$. Then (6.23) gives

$$\begin{aligned} b_{k+1} = a_{k+1}\eta^{-(k+1)} &\le (1-\gamma)a_k\eta^{-(k+1)} + Q_2 \frac{Q_7+Q_8}{1-\theta}\eta^{-(k+1)} \sum_{j=0}^{k-1} b_j \eta^j \zeta^{k-j} \\ &\le (1-\gamma)b_k\eta^{-1} + Q_2\frac{Q_7+Q_8}{1-\theta}\cdot\frac{1}{\eta}\sum_{j=0}^{k-1} Q_2\Big(\frac{\zeta}{\eta}\Big)^{k-j} \\ &\le Q_2\bigg((1-\gamma)\eta^{-1} + Q_2\frac{Q_7+Q_8}{1-\theta}\cdot\frac{1}{\eta}\cdot\frac{\frac{\zeta}{\eta}}{1-\frac{\zeta}{\eta}}\bigg) < Q_2, \end{aligned}$$

where the final inequality uses (6.24). It follows by induction that $b_k \le Q_2$ for every $k \in \mathbb{N}$, and thus $a_k \le Q_2\eta^k$ for every k. In particular, this gives

$$\Lambda_n(\mathcal{A}^k, \varphi) \le Q_2 \eta^k e^{nP(\varphi)} \tag{6.25}$$

for every $0 \le kT \le n$. It remains to relate $\mathcal{E}_n$ to $\mathcal{A}_n^{k_n}$ for some k_n.

Given $n \in \mathbb{N}$ large, choose $k_n \in [\frac{n}{3T}, \frac{n}{2T}] \cap \mathbb{N}$. We claim that

$$\mathcal{E}_n \subset \mathcal{A}_n^{k_n} \cup \bigg(\bigcup_{m=0}^{k_nT} \mathcal{L}_m(\mathcal{G}^- \setminus \mathcal{C}^p\mathcal{G})_{n-m}\bigg). \tag{6.26}$$

Indeed, given $w \in \mathcal{G}_n$, suppose that w is not contained in the right-hand side of (6.26); that is, $w \notin \mathcal{A}_n^{k_n}$ and $w_{(m,|w|]} \in \mathcal{C}^p\mathcal{G}$ for every $0 \le m \le k_nT$. Then by the definition of $\mathcal{A}_n^{k_n}$, there are $j < k$ and $i \in G_{\mathrm{s}}^M(w) \cap I_j$ such that $w_{(i,i']} \notin \mathcal{C}$ for any $i' \ge i + M - L$. Since $w_{(i,|w|]} \in \mathcal{C}\mathcal{G}$, this implies that there is $\ell \in [i, i+M-L]$ such that $w_{(\ell,|w|]} \in \mathcal{G}$. Applying **[III$_\mathrm{b}$]** to $w_{(i,i+M]}$ and $w_{(\ell,|w|]}$, we conclude that $w_{(i,|w|]} \in \mathcal{G}$, and hence $i - |c| \in S(w)$.[24] In particular, this proves that $S(w) \ne \emptyset$, so $w \notin \mathcal{E}_n$, establishing (6.26).

Using (6.26) together with (6.13) and (6.25), we now have the estimate

$$\begin{aligned} \Lambda_n(\mathcal{E},\varphi) &\le \Lambda_n(\mathcal{A}^{k_n},\varphi) + \sum_{m=0}^{k_nT} \Lambda_m(\mathcal{L},\varphi)\Lambda_{n-m}(\mathcal{G}^- \setminus \mathcal{C}^p\mathcal{G},\varphi) \\ &\le Q_2\eta^{k_n}e^{nP(\varphi)} + \sum_{m=0}^{k_nT} Q_2 e^{mP(\varphi)} Q_7 e^{(n-m)(P(\varphi)-\xi)} \\ &\le Q_2 e^{nP(\varphi)}\bigg(\eta^{\frac{n}{3T}} + Q_7 \sum_{\ell=n-k_nT}^{\infty} e^{-\ell\xi}\bigg) \\ &\le Q_2 e^{nP(\varphi)}\big(e^{\frac{1}{3T}(\log\eta)n} + Q_7(1-e^{-\xi})^{-1}e^{-\frac{n}{2}\xi}\big), \end{aligned}$$

and sending $n \to \infty$ gives

$$P(\mathcal{E},\varphi) \le P(\varphi) + \max(\tfrac{1}{3T}\log\eta, -\tfrac{\xi}{2}) < P(\varphi).$$

This completes the proof of Proposition 6.3, and hence of Theorem 3.2, modulo the proofs of Lemmas 6.8–6.11, which we give now in §§6.2.3–6.2.6.

---

[24]This is the only place in the paper where we use **[III$_\mathrm{b}$]**.

6.2.3. *Proof of Lemma 6.8.* We prove the lemma for $\mathcal{G}^+\setminus\mathcal{G}\mathcal{C}^s$; the other claim follows from a symmetric argument. Fix $\alpha > 0$ such that $\alpha(P(\varphi)+\|\varphi\|) < \varepsilon$, where $\varepsilon > 0$ is such that $P(\mathcal{C},\varphi) + \varepsilon < P(\varphi)$, so that there is $C > 0$ with $\Lambda_n(\mathcal{C},\varphi) \le Ce^{nP(\varphi)}$ for every n. Let N, Q_3 be as in Lemma 4.5, so that given any $n \in \mathbb{N}$ we can choose $m_n \in [\alpha n - \tau - N, \alpha n - \tau]$ for which $\Lambda_{m_n}(\mathcal{G},\varphi) \ge Q_3 e^{m_n P(\varphi)}$. Note that $m_n + \tau \le \alpha n$ and $\frac{1}{n} m_n \to \alpha$.

Given $w \in \mathcal{G}^+$, let $x \in \mathcal{L}$ be such that $wx \in \mathcal{G}$. For every $v \in \mathcal{G}$, it follows from **[I]** that there is $u = u(v,w) \in \mathcal{L}_{\le\tau}$ such that $vuwx \in \mathcal{G}$. From now on we will consider x as a function of w, and u as a function of w, v. Consider the collection $\mathcal{R} \subset \mathcal{G}^+$ given by

$$\mathcal{R}_n := \{w \in \mathcal{G}_n^+ \mid \text{ for all } v \in \mathcal{G}_{m_n} \text{ there is } k \ge 1 \text{ such that } vu(w_{[1,k]}) \in \mathcal{C}\}.$$

In particular, for every $w \in \mathcal{G}_n^+ \setminus \mathcal{R}$, there are $v \in \mathcal{G}_{m_n}$ and $u \in \mathcal{L}_{\le\tau}$ such that $vuw \in \mathcal{G}^+$ and $vu(w_{[1,k]}) \notin \mathcal{C}$ for every $1 \le k \le |w|$. We will consider v, u as functions of w whenever $w \in \mathcal{G}^+ \setminus \mathcal{R}$. Let

$$\mathcal{S} := \{w \in \mathcal{G}^+ \setminus \mathcal{R} \mid \text{there is } 1 \le j \le |vu| + L \text{ such that } (vuw)_{(j,|vuw|]} \in \mathcal{C}\}.$$

We will demonstrate below that $P(\mathcal{R}\cup\mathcal{S},\varphi) < P(\varphi)$. First we observe that if $w \in \mathcal{G}^+ \setminus (\mathcal{R}\cup\mathcal{S})$, then we have $vuw \notin \mathcal{C}$ and hence $vuw \in \mathcal{C}^p\mathcal{G}\mathcal{C}^s$, so that in particular there are $1 \le i \le j \le |vuw|$ such that

$$(vuw)_{[1,i]} \in \mathcal{C}^p, \qquad (vuw)_{(i,j]} \in \mathcal{G}, \qquad (vuw)_{(j,|vuw|]} \in \mathcal{C}^s.$$

By the choice of v and u, we have $i \le |vu|$, and by the definition of $\mathcal{S}$, we have $j > |vu| + L$. In particular, writing $\ell = j - |vu|$, we see that $w_{(\ell,|w|]} = (vuw)_{(j,|vuw|]} \in \mathcal{C}^s$, and also $w_{[1,\ell]}$ is the intersection of the two words $(vuwx)_{(i,j]} \in \mathcal{G}$ and $(vuwx)_{(|vu|,|vuwx|]} = wx \in \mathcal{G}$. Since $\ell \ge L$ this gives $w_{[1,\ell]} \in \mathcal{G}$ and thus $w \in \mathcal{G}\mathcal{C}^s$.

Having proved that $\mathcal{G}^+ \setminus \mathcal{G}\mathcal{C}^s \subset \mathcal{R}\cup\mathcal{S}$, it remains to estimate $P(\mathcal{R},\varphi)$ and $P(\mathcal{S},\varphi)$. To estimate $\Lambda_n(\mathcal{R},\varphi)$, we will estimate the partition sum of the collection $\{vuw \mid v \in \mathcal{G}_{m_n}, w \in \mathcal{R}_n\}$ in two different ways. First, note that every such vuw has length between $n + m_n$ and $n + m_n + \tau$, so along the same lines as in the proof of Lemma 4.5, we have

$$\begin{aligned}(6.27)\quad \sum_{v\in\mathcal{G}_{m_n}} \sum_{w\in\mathcal{R}_n} e^{\hat\varphi(vuw)} &\ge \sum_{v\in\mathcal{G}_{m_n}} \sum_{w\in\mathcal{R}_n} e^{\hat\varphi(v)} e^{\hat\varphi(w)} e^{-(\tau\|\varphi\|+|\varphi|_d)} \\ &\ge e^{-(\tau\|\varphi\|+|\varphi|_d)} \Lambda_{m_n}(\mathcal{G},\varphi)\Lambda_n(\mathcal{R},\varphi) \ge e^{-(\tau\|\varphi\|+|\varphi|_d)} Q_3 e^{m_n P(\varphi)} \Lambda_n(\mathcal{R},\varphi).\end{aligned}$$

On the other hand, given $v \in \mathcal{R}_n$ and $w \in \mathcal{G}_{m_n}$, we have $vuw \in \mathcal{C}_{m_n+|u|+k}\mathcal{L}_{n-k}$ for some $1 \le k \le n$, and thus as in (4.13),

$$\begin{aligned}(6.28)\quad \sum_{v\in\mathcal{G}_{m_n}} \sum_{w\in\mathcal{R}_n} e^{\hat\varphi(vuw)} &\le \sum_{k=1}^{n} \sum_{t=0}^{\tau} \sum_{x\in\mathcal{C}_{m_n+t+k}} \sum_{y\in\mathcal{L}_{n-k}} e^{\hat\varphi(x)} e^{\hat\varphi(y)} \\ &\le \sum_{k=1}^{n} \sum_{t=0}^{\tau} Ce^{(m_n+t+k)(P(\varphi)-\varepsilon)} Q_2 e^{(n-k)P(\varphi)}.\end{aligned}$$

Observe that for every $1 \leq k \leq n$ and $0 \leq t \leq \tau$ we have

$$\begin{aligned}(n-k)&P(\varphi) + (m_n + t + k)(P(\varphi) - \varepsilon) \\ &= (n + m_n + t)P(\varphi) - (m_n + t + k)\varepsilon \leq (n + m_n + \tau)P(\varphi) - m_n\varepsilon\end{aligned}$$

and so (6.28) gives

$$\sum_{v \in \mathcal{G}_{m_n}} \sum_{w \in \mathcal{R}_n} e^{\hat{\varphi}(vuw)} \leq n(\tau+1)Q_2 C e^{(n+m_n+\tau)P(\varphi)} e^{-m_n\varepsilon},$$

which together with (6.27) gives

$$\begin{aligned}\Lambda_n(\mathcal{R},\varphi) &\leq e^{\tau\|\varphi\| + |\varphi|_d} Q_3^{-1} e^{-m_n P(\varphi)} n(\tau+1) Q_2 C e^{(n+m_n+\tau)P(\varphi)} e^{-m_n\varepsilon} \\ &\leq e^{\tau\|\varphi\| + |\varphi|_d} Q_3^{-1} n(\tau+1) Q_2 C e^{(n+\tau)P(\varphi)} e^{-m_n\varepsilon},\end{aligned}$$

and we conclude that $P(\mathcal{R},\varphi) \leq P(\varphi) - \alpha\varepsilon < P(\varphi)$.

Now we consider $\Lambda_n(\mathcal{S},\varphi)$. For every $w \in \mathcal{S}_n$ we have $v \in \mathcal{G}_{m_n}$ and $u \in \mathcal{L}_t$ for some $t \leq \tau$ such that $vuw \in \mathcal{L}_j \mathcal{C}_{|vuw|-j}$ for some $j \leq |vu| + L$. This gives

$$\begin{aligned}\Lambda_n(\mathcal{S},\varphi) &= \sum_{w \in \mathcal{S}_n} e^{\hat{\varphi}(w)} \leq \sum_{w \in \mathcal{S}_n} e^{(\tau+m_n)\|\varphi\| + |\varphi|_d} e^{\hat{\varphi}(vuw)} \\ &\leq e^{\alpha n\|\varphi\| + |\varphi|_d} \sum_{t=0}^{\tau} \sum_{j=1}^{m_n+t+L} \Lambda_j(\mathcal{L},\varphi)\Lambda_{m_n+t+n-j}(\mathcal{C},\varphi).\end{aligned} \tag{6.29}$$

For each choice of t, j we get

$$\begin{aligned}\Lambda_j(\mathcal{L},\varphi)&\Lambda_{m_n+t+n-j}(\mathcal{C},\varphi) \leq Q_2 e^{jP(\varphi)} C e^{(m_n+t+n-j)(P(\varphi)-\varepsilon)} \\ &= Q_2 C e^{(m_n+t)P(\varphi)} e^{nP(\varphi)} e^{-\varepsilon(n+m_n+t-j)} \leq Q_2 C e^{\alpha n P(\varphi)} e^{nP(\varphi)} e^{-\varepsilon\ell}\end{aligned}$$

where $\ell = n + m_n + t - j \geq n - L$, and so

$$\sum_{j=1}^{m_n+t+L} \Lambda_j(\mathcal{L},\varphi)\Lambda_{m_n+t+n-j}(\mathcal{C},\varphi) \leq Q_2 C e^{\alpha n P(\varphi)} e^{nP(\varphi)} e^{-\varepsilon(n-L)} (1 - e^{-\varepsilon})^{-1}.$$

Together with (6.29), this gives

$$\Lambda_n(\mathcal{S},\varphi) \leq e^{\alpha n(P(\varphi) + \|\varphi\|)} e^{|\varphi|_d} Q_2 C e^{nP(\varphi)} (\tau+1) e^{-\varepsilon(n-L)} (1 - e^{-\varepsilon})^{-1},$$

and we conclude that $P(\mathcal{S},\varphi) \leq P(\varphi) + \alpha(P(\varphi) + \|\varphi\|) - \varepsilon < P(\varphi)$, where the last inequality uses our choice of α. This proves Lemma 6.8.

6.2.4. *Proof of Lemma 6.9.* We first describe the values of M that we will use, then give a computation that proves (6.14) (and shows that γ is independent of M). By Lemma 4.5 there are arbitrarily large values of m such that $\Lambda_m(\mathcal{G},\varphi) \geq Q_3 e^{mP(\varphi)}$. Given such an m, define $\pi\colon \mathcal{G}_m \to \bigsqcup_{i=0}^{\tau} s\mathcal{L} \cap \mathcal{G}_{|s|+m+i}$

using **[I]** by $\pi(w) = suw$ where $u = u(w) \in \mathcal{L}_{\leq\tau}$, and s is from the synchronising triple. Then we have

$$\sum_{i=0}^{\tau} \Lambda_{|s|+m+i}(s\mathcal{L} \cap \mathcal{G}, \varphi) \geq \sum_{w \in \mathcal{G}_m} e^{\hat{\varphi}(suw)} \geq \sum_{w \in \mathcal{G}_m} e^{\hat{\varphi}(su)} e^{\hat{\varphi}(w)} e^{-|\varphi|_d}$$
$$\geq e^{-(|s|+\tau)\|\varphi\| - |\varphi|_d} \Lambda_m(\mathcal{G}, \varphi),$$

so in particular there is $M \in [|s| + m, |s| + m + \tau]$ such that

$$\Lambda_M(s\mathcal{L} \cap \mathcal{G}, \varphi) \geq (\tau + 1)^{-1} e^{-(|s|+\tau)\|\varphi\| - |\varphi|_d} Q_3 e^{mP(\varphi)}$$
$$\geq (\tau + 1)^{-1} e^{-(|s|+\tau)(\|\varphi\| + P(\varphi)) - |\varphi|_d} Q_3 e^{MP(\varphi)}.$$

Putting $Q_9 = (\tau+1)^{-1} e^{-(|s|+\tau)(\|\varphi\|+P(\varphi)) - |\varphi|_d} Q_3$, we will use this in the form

$$\Lambda_M(s\mathcal{L} \cap \mathcal{G}, \varphi) \geq Q_9 e^{MP(\varphi)}. \tag{6.30}$$

Note that M can be taken arbitrarily large, and that Q_9 only depends on $Q_3, \tau, |s|, \varphi$.

Now we fix $v \in \mathcal{G}$ and $n \geq |v| + \tau + |rc| + M$. By **[I]** there is $p \in \mathcal{L}_{\leq\tau}$ such that $vpr \in \mathcal{G}$. Write $i = |vprc|$ and $m = n - i$; observe that for any $s' \in s\mathcal{L} \cap \mathcal{G}_M^+$ and $w \in \mathcal{H}_m^+(s')$, we have

$$vprcw \in \widetilde{\mathcal{H}_{\mathrm{s}}^M}(v, i) := \{x \in \mathcal{H}_n^+(vprc) \mid i \in G_{\mathrm{s}}^M(x)\},$$

where we use the synchronising property of (r, c, s). That is, i represents a 'candidate' good occurrence of the synchronising triple in $vprcw$. We first estimate how many such words $vprcw$ there are, and then show that in most of them, i must be a genuinely good occurrence; that is, there is typically not a long subword in $\mathcal{C}$ beginning in position i.

Proposition 4.6 gives[25]

$$\Lambda_m(\mathcal{H}^+(s'), \varphi) \geq Q_4^{-1} e^{(m-M)P(\varphi)} e^{\hat{\varphi}(s')} \tag{6.31}$$

so we have

$$\Lambda_n(\widetilde{\mathcal{H}_{\mathrm{s}}^M}(v, i), \varphi) \geq \sum_{s' \in s\mathcal{L} \cap \mathcal{G}_M} \sum_{w \in \mathcal{H}_m^+(s')} e^{\hat{\varphi}(vprcw)}$$
$$\geq \sum_{s' \in s\mathcal{L} \cap \mathcal{G}_M} \sum_{w \in \mathcal{H}_m^+(s')} e^{\hat{\varphi}(vprc)} e^{\hat{\varphi}(w)} e^{-|\varphi|_d}$$
$$\geq e^{\hat{\varphi}(v) - |prc|\|\varphi\| - 2|\varphi|_d} \sum_{s' \in s\mathcal{L} \cap \mathcal{G}_M} \Lambda_m(\mathcal{H}^+(s'), \varphi).$$

Using the bounds from (6.30) and (6.31), we get

$$\sum_{s' \in s\mathcal{L} \cap \mathcal{G}_M} \Lambda_m(\mathcal{H}^+(s'), \varphi) \geq \sum_{s' \in s\mathcal{L} \cap \mathcal{G}_M} Q_4^{-1} e^{(m-M)P(\varphi)} e^{\hat{\varphi}(s')}$$
$$\geq Q_4^{-1} e^{(m-M)P(\varphi)} \Lambda_M(s\mathcal{L} \cap \mathcal{G}, \varphi) \geq Q_4^{-1} e^{mP(\varphi)} Q_9,$$

[25] The statement of Proposition 4.6 is given for $\mathcal{H}_m(s', 1)$ instead of $\mathcal{H}^+$, but the proof constructs words in $\mathcal{G}^+$ and so in fact proves the bound for $\mathcal{H}^+(s')$.

which gives

$$\Lambda_n(\widetilde{\mathcal{H}_{\mathrm{s}}^M}(v,i),\varphi) \geq e^{\hat\varphi(v)} e^{mP(\varphi)} e^{-|prc|\|\varphi\|-2|\varphi|_d} Q_4^{-1} Q_9. \tag{6.32}$$

The upper bound in Proposition 4.6 gives

$$\Lambda_n(\mathcal{H}^+(v),\varphi) \leq Q_4 e^{(n-|v|)P(\varphi)} e^{\hat\varphi(v)} \leq Q_4 e^{(\tau+|rc|)P(\varphi)} e^{mP(\varphi)} e^{\hat\varphi(v)}; \tag{6.33}$$

together with (6.32) this gives

$$\Lambda_n(\widetilde{\mathcal{H}_{\mathrm{s}}^M}(v,i),\varphi) \geq 2\gamma\Lambda_n(\mathcal{H}^+(v),\varphi) \tag{6.34}$$

for $\gamma = \frac{1}{2} e^{-(\tau+|rc|)(P(\varphi)+\|\varphi\|)-2|\varphi|_d} Q_4^{-2} Q_9$. Now we must exclude those words $x \in \widetilde{\mathcal{H}_{\mathrm{s}}^M}(v,i)$ for which $x_{(i,i']} \in \mathcal{C}$ for some $i' \in [i+M-L,n]$. Let $\mathcal{B}(v,i)$ be the set of such x; note that $\mathcal{B}(v,i) \subset \bigcup_{i'=i+M-L}^n vprc\mathcal{C}_{i'-i}\mathcal{L}_{n-i'}$, so taking $C,\varepsilon > 0$ such that $\Lambda_k(\mathcal{C},\varphi) \leq Ce^{k(P(\varphi)-\varepsilon)}$ for all $k \in \mathbb{N}$, we have

$$\begin{aligned}
\Lambda_n(\mathcal{B}(v,i)) &\leq \sum_{i'=i+M-L}^n e^{\hat\varphi(vprc)} \Lambda_{i'-i}(\mathcal{C},\varphi)\Lambda_{n-i'}(\mathcal{L},\varphi) \\
&\leq e^{\hat\varphi(v)} e^{\tau\|\varphi\|+\hat\varphi(rc)} \sum_{i'=i+M-L}^n Ce^{(i'-i)(P(\varphi)-\varepsilon)} Q_2 e^{(n-i')P(\varphi)} \\
&\leq e^{\hat\varphi(v)} e^{\tau\|\varphi\|+\hat\varphi(rc)} e^{(n-i)P(\varphi)} CQ_2 e^{-(M-L)\varepsilon}(1-e^{-\varepsilon})^{-1}.
\end{aligned}$$

Since $e^{(n-i)P(\varphi)} = e^{(m+|c|)P(\varphi)}$, by (6.33) we can take M large enough that

$$\Lambda_n(\mathcal{B}(v,i)) \leq \gamma\Lambda_n(\mathcal{H}^+(v),\varphi)$$

for every n and v, which together with (6.34) shows that

$$\Lambda_n(\widetilde{\mathcal{H}_{\mathrm{s}}^M}(v,i) \setminus \mathcal{B}(v,i),\varphi) \geq \gamma\Lambda_n(\mathcal{H}^+(v),\varphi).$$

Finally, $\widetilde{\mathcal{H}_{\mathrm{s}}^M}(v,i) \setminus \mathcal{B}(v,i) \subset \mathcal{H}_{\mathrm{s}}^M(v)$, completing the proof of Lemma 6.9.

6.2.5. *Proof of Lemma 6.10.* Given $\beta,\delta > 0$ and $w \in \hat{\mathcal{C}}_n^{\beta,\delta}$, let i_a, i_a' be as in the definition of $\hat{\mathcal{C}}$. Write $I_a = (i_a, i_a'] \cap \mathbb{N}$ and observe that if some $i \in [1,|w|]$ is contained in I_a for three distinct choices of a, then one of the corresponding intervals I_a is contained in the union of the other two, and hence can be removed from the collection without changing $\bigcup_a I_a$. Thus without loss of generality we may assume that every $i \in [1,|w|]$ is contained in at most two of the I_a, and by re-indexing if necessary, we have

$$i_1 < i_1' \leq i_3 < i_3' \leq \cdots, \qquad i_2 < i_2' \leq i_4 < i_4' \leq \cdots.$$

Either $\sum_{a \text{ even}}(i_a' - i_a) \geq \delta n$ or $\sum_{a \text{ odd}}(i_a' - i_a) \geq \delta n$. Write j_b, j_b' for the indices in the larger sum, so $b = 1,\dots,B'$ where $B' \leq \beta n$. We see that $w \in \mathcal{L}_{j_1}\mathcal{C}_{j_1'-j_1}\mathcal{L}_{j_2-j_1'}\mathcal{C}_{j_2'-j_2}\cdots\mathcal{C}_{j_b'-j_b}\mathcal{L}_{n-j_b'}$; in particular,

$$\hat{\mathcal{C}}_n^{\beta,\delta} \subset \bigcup_{B'=1}^{\lfloor\beta n\rfloor} \bigcup_{\vec{j},\vec{j}'} \mathcal{L}_{j_1}\mathcal{C}_{j_1'-j_1}\mathcal{L}_{j_2-j_1'}\mathcal{C}_{j_2'-j_2}\cdots\mathcal{C}_{j_b'-j_b}\mathcal{L}_{n-j_b'},$$

where the inner union is over all sequences $0 \le j_1 < j_1' \le j_2 \cdots \le j_{B'}' \le n$. Assume that $\beta < \frac{1}{2}$, so $2B' \le 2\lfloor \beta n \rfloor \le n$; then the number of such sequences is at most[26] $\binom{2(n+1)}{2B'} \le \binom{2n+2}{2\lfloor \beta n \rfloor} \le 4\binom{2n}{2\lfloor \beta n \rfloor} 4(2n+1)e^{h(\beta)2n+1}$. For each such sequence we can write $\ell(\vec{j}, \vec{j}') = \sum_b (j_b' - j_b) \ge \delta n$ and get (as in (4.13))

$$\begin{aligned}\Lambda_n(&\mathcal{L}_{j_1}\mathcal{C}_{j_1'-j_1}\mathcal{L}_{j_2-j_1'}\mathcal{C}_{j_2'-j_2}\cdots\mathcal{C}_{j_b'-j_b}\mathcal{L}_{n-j_b'})\\ &\le Q_2^{B'+1}e^{(n-\ell(\vec{j},\vec{j}'))P(\varphi)}C^{B'}e^{\ell(\vec{j},\vec{j}')(P(\varphi)-\varepsilon)} \le Q_2(Q_2C)^{\beta n}e^{nP(\varphi)}e^{-\delta\varepsilon n}.\end{aligned}$$

Summing over all choices of B' and of $\vec{j}, \vec{j}'$ gives

$$\Lambda_n(\hat{\mathcal{C}}^{\beta,\delta}) \le 4\beta n(2n+1)e^{h(\beta)2n+1}Q_2 e^{(\log(Q_2C))\beta n}e^{nP(\varphi)}e^{-\delta\varepsilon n},$$

and we conclude that

$$P(\hat{\mathcal{C}}^{\beta,\delta}, \varphi) \le 2h(\beta) + \beta\log(Q_2C) + P(\varphi) - \delta\varepsilon.$$

For small enough β this is $< P(\varphi)$, which proves Lemma 6.10.

6.2.6. *Proof of Lemma 6.11.* First we prove the inclusions in (6.19), then we prove the estimate in (6.20). For the first inclusion in (6.19), if $v \in \mathcal{Z}^k$ and $w \in \mathcal{H}^+(v)$, then $w_{[1,kT]} = v_{[1,kT]} \in \mathcal{A}^k$ by the definition of $\mathcal{Z}^k$, so $w \in \mathcal{A}^k$.

The second inclusion requires more work. As in the lemma, write $m_k = (k+\frac{1}{2})T - (\tau + |rc|)$. If $w_{[1,m_k]} \notin \mathcal{GC}^s$, then $w \in \mathcal{Y}^k$, so we may assume that $w_{[1,m_k]} \in \mathcal{GC}^s$. Let $i \in [1, m_k]$ be maximal such that $w_{[1,i]} \in \mathcal{G}$; it follows that $w_{(i,m_k]} \in \mathcal{C}$. Let $k' = \lfloor \frac{i}{T} \rfloor$, so that $0 \le k' \le k$, and let $j \le k'$ be maximal such that $w_{[1,jT]} \in \mathcal{A}^j$.

First consider the case when $j = k$. Then we must have $k' = k$ and hence $i \ge kT$. Taking the minimal $i \ge kT$ such that $v := w_{[1,i]} \in \mathcal{G}$, we see that $v \in \mathcal{Z}^k$ and $w \in \mathcal{H}^+(v)$. So we turn our attention to the case $j < k$.

Lemma 6.12. *There is $\ell \in [k'T, |w|]$ such that $w_{(jT,\ell]}$ contains $B = k' - j$ intervals $\{(i_a, i_a']\}_{a=1}^B$ such that*

$$w_{(i_a,i_a']} \in \mathcal{C} \text{ for every } a, \text{ and } \#\bigcup_{a=1}^B (i_a, i_a'] \ge \tfrac{1}{2}(\ell - jT). \tag{6.35}$$

Proof. If $j = k'$ then taking $\ell = k'T$ suffices. So assume $j < k'$. Given $1 \le a \le B = k' - j$, we have $w_{[1,(j+a)T]} \notin \mathcal{A}^{j+a}$ by maximality of j. Thus there are $j_a < j + a$ and $i_a \in I_{j_a}$ such that $w_{(i_a,i']} \notin \mathcal{C}$ for all $i' \in [i_a + M - L, (j+a)T]$. Note that $j \le j_a$ because $w_{[1,jT]} \in \mathcal{A}^j$, and so $jT < i_a \le (j + a - \frac{1}{2})T$.

On the other hand, $w \in \mathcal{A}^k$ and so for each a there is $i_a' \in [i + M - L, |w|]$ such that $w_{(i_a,i_a']} \in \mathcal{C}$. By the previous paragraph we must have $i_a' > (j+a)T$. Let $\ell = \max_a i_a'$. Because $i_a \le k'T$ for all $a = 1, \dots, B$, we have

$$\bigcup_{a=1}^B (i_a, i_a'] \supset \left(\bigcup_{a=1}^B (j + a - \tfrac{1}{2})T, (j+a)T]\right) \cup (k'T, \ell],$$

which proves (6.35). ∎

[26]The factor of 2 in the top half comes since we allow $j_b' = j_{b+1}$; to associate each such sequence to a strictly increasing sequence we can duplicate each of the numbers $0, 1, \dots, n$.

Let ℓ be given by Lemma 6.12. If $\ell \geq kT$ then we have $w \in \mathcal{Y}^k$, since $w_{(jT,\ell]} \in \hat{\mathcal{C}}^{\beta,\delta}$ for $\delta = \frac{1}{8}$ and $\beta \geq \frac{1}{2T}$, by the estimate $B = k' - j \leq \frac{1}{T}(\ell - jT)$.

Now consider the case $\ell < kT$; we claim that in this case we have $w_{(jT,m_k]} \in \hat{\mathcal{C}}^{\beta,\delta}$. Put $i_0 = i$ and $i_0' = m_k$; consider the collection of intervals $\{(i_a, i_a']\}_{a=0}^B \subset (jT, m_k]$. Note that $k' < k$ since $k'T \leq \ell$, and so

$$B + 1 = k' - j + 1 \leq k - j \leq \tfrac{1}{T}(m_k - jT).$$

It remains to show that $\#\bigcup_{a=0}^B (i_a, i_a'] \geq \frac{1}{4}(m_k - jT)$. From (6.35) we have $\#\bigcup_{a=1}^B (i_a, i_a'] \geq \frac{1}{2}(\ell - jT)$. Moveover, this union is contained in $(jT, \ell]$, so writing $\ell' = \max(i, \ell)$, we use the general inequality $\frac{w+x}{y+z} \geq \min(\frac{w}{y}, \frac{x}{z})$ to get

$$\frac{\#\bigcup_{a=0}^B (i_a, i_a']}{m_k - jT} = \frac{(\#\bigcup_{a=1}^B (i_a, i_a']) + \#(\ell', m_k]}{(\ell - jT) + (m_k - \ell)} \geq \min\left(\frac{1}{2}, \frac{m_k - \ell'}{m_k - \ell}\right).$$

We have $\ell' - \ell \leq T$ and $m_k - \ell' \geq \frac{T}{2} - (\tau + |rc|) = M \geq \frac{T}{3}$ (using the bound following (6.16)), so this gives

$$\frac{m_k - \ell'}{m_k - \ell} = \frac{m_k - \ell'}{(m_k - \ell') + (\ell' - \ell)} = \frac{1}{1 + \frac{\ell' - \ell}{m_k - \ell'}} \geq \frac{1}{1 + \frac{T}{T/3}} = \frac{1}{4}.$$

This proves that $w_{(jT,m_k]} \in \hat{\mathcal{C}}^{\beta,\delta}$, and to complete the proof of Lemma 6.11 we must prove the estimate in (6.20). We do this using (6.15)–(6.17), which give (by (4.13))

$$\begin{aligned}
\Lambda_n((\mathcal{G}^+ \setminus (\mathcal{G}\mathcal{C}^s))_{m_k} \mathcal{L}_{n-m_k}, \varphi) &\leq Q_7 e^{m_k(P(\varphi) - \xi)} Q_2 e^{(n-m_k)P(\varphi)} \\
&\leq Q_7 Q_2 e^{nP(\varphi)} e^{-m_k \xi} \leq Q_7 Q_2 \zeta^k e^{nP(\varphi)}
\end{aligned}$$

for the first part of $\mathcal{Y}^k$, and for the second part,

$$\begin{aligned}
\Lambda_n\bigg(\bigcup_{j=0}^{k-1} \bigcup_{\ell=kT}^{n} \mathcal{A}^j_{jT} \hat{\mathcal{C}}^{\beta,\delta}_{\ell-jT} \mathcal{L}_{n-\ell}, \varphi\bigg) &\leq \sum_{j=0}^{k-1} \sum_{\ell=kT}^{n} \Lambda_{jT}(\mathcal{A}^j, \varphi) \Lambda_{\ell-jT}(\hat{\mathcal{C}}^{\beta,\delta}, \varphi) \Lambda_{n-\ell}(\mathcal{L}, \varphi) \\
&\leq \sum_{j=0}^{k-1} \sum_{\ell=kT}^{n} \Lambda_{jT}(\mathcal{A}^j, \varphi) Q_8 \theta^{\ell-jT} e^{(\ell-jT)P(\varphi)} Q_2 e^{(n-\ell)P(\varphi)} \\
&\leq Q_8 Q_2 \sum_{j=0}^{k-1} \Lambda_{jT}(\mathcal{A}^j, \varphi) \frac{\theta^{(kT-jT)}}{1-\theta} e^{(n-jT)P(\varphi)}.
\end{aligned}$$

Adding the two estimates (and loosening the first) gives (6.20).[27] This proves Lemma 6.11.

[27]Strictly speaking, in (6.20) we could put Q_7 outside the fraction, and only multiply it by the term in the sum corresponding to $j = 0$, but the looser estimate leads to less cumbersome bookkeeping and is sufficient for our purposes.

7. Proofs of other results

In this section we give the remaining proofs. First we prove Theorem 1.3, then the characterisation of strong positive recurrence in §2.1.3, then the factor results Propositions 3.16 and 3.17, and we finish by proving the claims in Remark 3.3.

Proof of Theorem 1.3. Before giving the proof of Theorem 1.3, we prove a lemma that we will use several times below. Given $\mathcal{A} \subset \mathcal{L}$, let

$$\hat{P}(\mathcal{A}, \varphi) := \sup_{n\in\mathbb{N}} \frac{1}{n} \log \Lambda_n(\mathcal{A}, \varphi)$$

and consider the collection $\mathcal{A}^* = \{w^1 \cdots w^k \in \mathcal{L} \mid w^i \in \mathcal{A} \text{ for all } i\}$.

Lemma 7.1. *Let φ be Hölder and $\mathcal{A} \subset \mathcal{L}_{\geq M}$ for some $M \in \mathbb{N}$. Then*

$$P(\mathcal{A}^*, \varphi) \leq \hat{P}(\mathcal{A}, \varphi) + h(\tfrac{1}{M}), \tag{7.1}$$

where $h(\delta) = -\delta \log \delta - (1-\delta)\log(1-\delta)$.

Proof. Given $n, k \in \mathbb{N}$, let $\mathbf{N}_k = \{(n_1, n_2, \dots, n_k) \in \mathbb{N}^k \mid \sum n_i = n \text{ and } n_i \geq M \text{ for all } 1 \leq i \leq k\}$. Note that $\mathbf{N}_k$ is empty for all $k > \frac{n}{M}$. Now we have

$$\begin{aligned}
\Lambda_n(\mathcal{A}^*, \varphi) &= \sum_{k\in\mathbb{N}} \sum_{\mathbf{N}_k} \sum_{\substack{w^1 \in \mathcal{A}_{n_1}, \cdots w^k \in \mathcal{A}_{n_k} \\ w^1 \cdots w^k \in \mathcal{L}}} e^{\hat{\varphi}(w^1 \cdots w^k)} \\
&\leq \sum_{k} \sum_{\mathbf{N}_k} \sum_{w^1 \in \mathcal{A}_{n_1}, \cdots w^k \in \mathcal{A}_{n_k}} e^{\sum_{i=1}^k \hat{\varphi}(w^i)} \\
&\leq \sum_{k} \sum_{\mathbf{N}_k} \prod_{i=1}^{k} \Lambda_{n_i}(\mathcal{A}, \varphi) \leq \sum_{k=1}^{\lfloor n/M \rfloor} (\#\mathbf{N}_k) e^{n\hat{P}(\mathcal{A},\varphi)}
\end{aligned}$$

Recalling Lemma 5.8, we see that for every $k \leq \frac{n}{M}$ we have

$$\#\mathbf{N}_k \leq \tbinom{n}{k} \leq (n+1) e^{h(\frac{1}{M})n+1},$$

which gives the estimate

$$\Lambda_n(\mathcal{A}^*, \varphi) \leq \tfrac{n}{M}(n+1) e^{h(\frac{1}{M})n+1} e^{n\hat{P}(\mathcal{A},\varphi) + \frac{n}{M}|\varphi|_d}.$$

We conclude that $P(\mathcal{A}^*, \varphi) \leq \hat{P}(\mathcal{A}, \varphi) + h(\frac{1}{M})$, as claimed. ∎

Now we proceed with the proof of Theorem 1.3. As in the statement of the theorem, let $\mathcal{C}^{\pm}$ satisfy (1.2) and [$\mathbf{I}^*$], and fix $\varepsilon > 0$. We start by choosing parameters M, τ, N to satisfy certain pressure estimates; then we use these to define $\mathcal{C}^p, \mathcal{C}^s$. Finally, we define $\mathcal{G}$ and verify [**I**], [**III**].

Step 1 (*choosing* M, τ, N): Choose $M \in \mathbb{N}$ such that $h(\frac{1}{M}) < \varepsilon$ and

$$\hat{P}(\mathcal{C}^-_{\geq M}, \varphi) < P(\mathcal{C}^-, \varphi) + \varepsilon, \qquad \hat{P}(\mathcal{C}^+_{\geq M}, \varphi) < P(\mathcal{C}^+, \varphi) + \varepsilon. \tag{7.2}$$

Let $\tau = \tau(M)$ be as in **[I*]** and consider the collections

$$\mathcal{D}^- := \{w \in \mathcal{L} \mid \text{there exists } x \in \mathcal{L}_{\leq \tau+M} \text{ such that } wx \in \mathcal{C}^-\},$$
$$\mathcal{D}^+ := \{w \in \mathcal{L} \mid \text{there exists } x \in \mathcal{L}_{\leq \tau+M} \text{ such that } xw \in \mathcal{C}^+\}.$$

These have the same pressures as $\mathcal{C}^\pm$; indeed,

$$\Lambda_n(\mathcal{D}^-, \varphi) = \sum_{i=0}^{\tau+M} \sum_{\substack{w \in \mathcal{D}_n^- \\ |x(w)|=i}} e^{\hat\varphi(w)} \leq \sum_{i=0}^{\tau+M} \sum_{v \in \mathcal{C}_{n+i}^-} e^{\hat\varphi(v)+|\varphi|_d+i\|\varphi\|}$$
$$\leq (\tau+M+1)e^{|\varphi|_d+(\tau+M)\|\varphi\|} e^{(n+\tau+M)\hat P(\mathcal{C}_{\geq n}^-,\varphi)}.$$

Sending $n \to \infty$ gives $P(\mathcal{D}^-, \varphi) \leq P(\mathcal{C}^-, \varphi)$, and the estimate for $\mathcal{D}^+$ is similar. Thus we can choose N large enough that $\frac{\log 2}{N} < \varepsilon$ and

$$\hat P(\mathcal{D}_{\geq N}^-, \varphi) < P(\mathcal{C}^-, \varphi) + \varepsilon, \qquad \hat P(\mathcal{D}_{\geq N}^+, \varphi) < P(\mathcal{C}^+, \varphi) + \varepsilon. \tag{7.3}$$

Step 2 (*definition of* $\mathcal{C}^p, \mathcal{C}^s$): Consider the collections

$$\mathcal{C}^p := (\mathcal{C}_{\geq M}^- \cup \mathcal{D}_{\geq N}^+)^*, \qquad \mathcal{C}^s := (\mathcal{C}_{\geq M}^+ \cup \mathcal{D}_{\geq N}^-)^*.$$

By Lemma 7.1 we have

$$P(\mathcal{C}^p, \varphi) \leq \hat P(\mathcal{C}_{\geq M}^- \cup \mathcal{D}_{\geq N}^+, \varphi) + h(\tfrac{1}{M}). \tag{7.4}$$

To estimate the $\hat P$ term we note that for $n \in [M, N)$ we have

$$\Lambda_n(\mathcal{C}_{\geq M}^- \cup \mathcal{D}_{\geq N}^+, \varphi) = \Lambda_n(\mathcal{C}_{\geq M}^-, \varphi) \leq e^{n\hat P(\mathcal{C}_{\geq M}^-,\varphi)} \leq e^{n(P(\mathcal{C}^-,\varphi)+\varepsilon)}$$

using (7.2), while for $n \geq N$ we have

$$\Lambda_n(\mathcal{C}_{\geq M}^- \cup \mathcal{D}_{\geq N}^+, \varphi) \leq \Lambda_n(\mathcal{C}_{\geq M}^-, \varphi) + \Lambda_n(\mathcal{D}_{\geq N}^+, \varphi)$$
$$\leq 2e^{n\max\{\hat P(\mathcal{C}_{\geq M}^-,\varphi), \hat P(\mathcal{D}_{\geq N}^+,\varphi)\}} \leq 2e^{n(\max\{P(\mathcal{C}^-,\varphi),P(\mathcal{C}^+,\varphi)\}+\varepsilon)}$$

using (7.2) and (7.3). We conclude that

$$\hat P(\mathcal{C}_{\geq M}^- \cup \mathcal{D}_{\geq N}^+, \varphi) \leq P(\mathcal{C}^- \cup \mathcal{C}^+, \varphi) + \varepsilon + \tfrac{\log 2}{N} < P(\mathcal{C}^- \cup \mathcal{C}^+, \varphi) + 2\varepsilon,$$

Together with (7.4) and the estimate on $h(\frac{1}{M})$ this gives

$$P(\mathcal{C}^p, \varphi) < P(\mathcal{C}^- \cup \mathcal{C}^+, \varphi) + 3\varepsilon. \tag{7.5}$$

The estimate for $P(\mathcal{C}^s, \varphi)$ is similar. Sending $\varepsilon \to 0$, the right-hand side can be made $< P(\varphi)$.

Step 3 (*definition of* $\mathcal{G}$): Now we describe $\mathcal{G}$ such that **[I]** and **[III]** hold and we have $P(\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s, \varphi) < P(\varphi)$. Let

$$\begin{aligned}\mathcal{G} := \{w \in \mathcal{L} \setminus (\mathcal{D}^+ \cup \mathcal{D}^-) \mid\ & w_{[1,i]} \notin \mathcal{C}^-, w_{(|w|-i,|w|]} \notin \mathcal{C}^+ \\ & \text{for all } i \geq M, \text{ and } w_{[1,i]} \notin \mathcal{D}^+, w_{(|w|-i,|w|]} \notin \mathcal{D}^- \text{ for all } i \geq N\}.\end{aligned} \tag{7.6}$$

Given $w \in \mathcal{L}$, decompose w as $w = u^p v u^s$ by beginning with $v = w$ and $u^p = u^s = \emptyset$, and then proceeding as follows.

(1) Choose the smallest $i \in [1,|v|]$ such that $v_{[1,i]} \in \mathcal{C}^-_{\geq M} \cup \mathcal{D}^+_{\geq N}$ (if such an i exists); then replace u^p with $u^p v_{[1,i]}$ and replace v with $v_{(i,|v|]}$. Iterate this step until no such i exists; note that $u^p \in \mathcal{C}^p$.
(2) Take the resulting word v and choose the smallest $i \in [1,|v|]$ such that $v_{(|v|-i,|v|]} \in \mathcal{C}^+_{\geq M} \cup \mathcal{D}^-_{\geq N}$ (if such an i exists); then replace u^s with $v_{(|v|-i,|v|]}u^s$ and v with $v_{[1,|v|-i]}$. Iterate this step until no such i exists; note that $u^s \in \mathcal{C}^s$.
(3) Observe that the resulting word v satisfies $v \in \mathcal{G} \cup \mathcal{D}^+ \cup \mathcal{D}^-$ by the definition of $\mathcal{G}$.

We conclude that $\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s \subset \mathcal{C}^p(\mathcal{D}^+ \cup \mathcal{D}^-)\mathcal{C}^s$, and in particular,[28]

$$P(\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s, \varphi) \leq \max\{P(\mathcal{C}^p,\varphi), P(\mathcal{D}^+,\varphi), P(\mathcal{D}^-,\varphi), P(\mathcal{C}^s,\varphi)\} < P(\varphi).$$

It remains to show that $\mathcal{G}$ satisfies **[I]** and **[III]**. For **[I]**, start by observing that $\mathcal{G} \subset \mathcal{G}^M(\mathcal{C}^\pm)$, and so by **[I*]**, for every $v,w \in \mathcal{G}$ there is $u \in \mathcal{L}_{\leq\tau}$ such that $vuw \in \mathcal{L}$. We claim that in fact $vuw \in \mathcal{G}$, which will establish **[I]**. To show this, we prove that

⟨**1**⟩ $vuw \notin \mathcal{D}^+$ and $vuw \notin \mathcal{D}^-$;
⟨**2**⟩ $(vuw)_{[1,i]} \notin \mathcal{C}^-$ and $(vuw)_{(|vuw|-i,|vuw|]} \notin \mathcal{C}^+$ for all $i \geq M$;
⟨**3**⟩ $(vuw)_{[1,i]} \notin \mathcal{D}^+$ and $(vuw)_{(|vuw|-i,|vuw|]} \notin \mathcal{D}^-$ for all $i \geq N$.

In each case we prove only the first assertion; the second follows by a symmetrical argument.

For ⟨**1**⟩, given any $x \in \mathcal{L}_{\leq\tau+M}$ we have $xv \notin \mathcal{C}^+$ since $v \notin \mathcal{D}^+$, and hence $xvuw \notin \mathcal{C}^+$ by (1.2); this proves $vuw \in \mathcal{D}^+$. For ⟨**2**⟩ we consider $(vuw)_{[1,i]}$ in the following three cases.

- $M \leq i \leq |v|$. Then $v \in \mathcal{G}$ gives $(vuw)_{[1,i]} = v_{[1,i]} \notin \mathcal{C}^-$.
- $|v| < i \leq |vu| + M$. Then since $v \notin \mathcal{D}^-$ and $i - |v| \leq \tau + M$, we must have $(vuw)_{[1,i]} = v(uw)_{[1,i-|v|]} \notin \mathcal{C}^-$.
- $i > |vu| + M$. Then $w \in \mathcal{G}$ gives $w_{[1,i-|vu|]} \notin \mathcal{C}^-$, so (1.2) gives $(vuw)_{[1,i]} \notin \mathcal{C}^-$.

For ⟨**3**⟩, given any $x \in \mathcal{L}_{\leq\tau+M}$ and $i \geq N$, there are two cases: either $|v| \leq i$, in which case $xv \notin \mathcal{C}^+$ (since $v \notin \mathcal{D}^+$) gives $x(vuw)_{[1,i]} \notin \mathcal{C}^+$ by (1.2); or $|v| > i$, in which case $x(vuw)_{[1,i]} = x(v_{[1,i]}) \notin \mathcal{C}^+$ since $v \in \mathcal{G}$, and thus $(vuw)_{[1,i]} \notin \mathcal{D}^+$. We conclude that $vuw \in \mathcal{G}$, which verifies **[I]**.

The proof of **[III]** has a similar flavour. If $uvw \in \mathcal{L}$ and $uv, vw \in \mathcal{G}$, we show that $uvw \in \mathcal{G}$, and that $v \in \mathcal{G}$ if $|v| \geq N$. As above, we verify the conditions involving $\mathcal{C}^-$ and $\mathcal{D}^+$; the conditions involving $\mathcal{C}^+$ and $\mathcal{D}^-$ follow from symmetric arguments. We start with uvw.

⟨**1**⟩ For $x \in \mathcal{L}_{\leq\tau+M}$, $uv \notin \mathcal{D}^+$ gives $xuv \notin \mathcal{C}^+$, so $xuvw \notin \mathcal{C}^+$ by (1.2), hence $uvw \notin \mathcal{D}^+$.
⟨**2**⟩ For $i \geq M$, we check $(uvw)_{[1,i]} \notin \mathcal{C}^-$ in the following three cases.
  - $M \leq i \leq |uv|$. Then $uv \in \mathcal{G}$ gives $(uvw)_{[1,i]} = (uv)_{[1,i]} \notin \mathcal{C}^-$.

[28]In fact, the estimates given here and earlier show that we can make $P(\mathcal{C}^p \cup \mathcal{C}^s \cup (\mathcal{L} \setminus \mathcal{C}^p\mathcal{G}\mathcal{C}^s),\varphi)$ as close to $P(\mathcal{C}^+ \cup \mathcal{C}^-,\varphi)$ as we like by taking M,N large.

- $|uv| < i \le |uv| + M$. Then $uv \notin \mathcal{D}^-$ implies that $(uvw)_{[1,i]} = (uv)w_{[1,i-|uv|]} \notin \mathcal{C}^-$ since $i - |uv| \le M$.
- $i > |uv| + M \ge |u| + M$. Then $vw \in \mathcal{G}$ gives $(vw)_{[1,i-|u|]} \notin \mathcal{C}^-$, hence $(uvw)_{[1,i]} \notin \mathcal{C}^-$ by (1.2).

$\langle \mathbf{3} \rangle$ Given any $x \in \mathcal{L}_{\le\tau+M}$ and $i \ge N$, there are two cases. If $|uv| \le i$, then $uv \notin \mathcal{D}^+$ gives $xuv \notin \mathcal{C}^+$ and hence $x(uvw)_{[1,i]} \notin \mathcal{C}^+$ by (1.2). If $|uv| \ge i$, then $uv \in \mathcal{G}$ gives $x(uvw)_{[1,i]} = x(uv)_{[1,i]} \notin \mathcal{C}^+$. In both cases we deduce that $(uvw)_{[1,i]} \notin \mathcal{D}^+$.

We conclude by showing that $v \in \mathcal{G}$ whenever $|v| \ge N$.

$\langle \mathbf{1} \rangle$ Given $x \in \mathcal{L}_{\le\tau+M}$, we have $xv = x(vw)_{[1,|v|]} \notin \mathcal{C}^+$ since $vw \in \mathcal{G}$ and $|v| \ge N$; thus $v \notin \mathcal{D}^+$.

$\langle \mathbf{2} \rangle$ Given $i \ge M$ we have $v_{[1,i]} = (vw)_{[1,i]} \notin \mathcal{C}^-$ since $vw \in \mathcal{G}$.

$\langle \mathbf{3} \rangle$ Given $i \ge N$ we have $v_{[1,i]} = (vw)_{[1,i]} \notin \mathcal{D}^+$ since $vw \in \mathcal{G}$.

This establishes **[III]** for $\mathcal{G}$ and completes the proof of Theorem 1.3.

Proof of claim in §2.1.3: $\Delta_a[\Phi] > 0$ iff (2.10). The definition of strong positive recurrence in [Sar01] involves positivity of a certain **discriminant** $\Delta_a[\Phi]$. More precisely, one defines the **induced pressure function** $\gamma(p) := P_G(\overline{\Phi + p})$ for $p \in \mathbb{R}$; this function can take finite or infinite values, and we will not need to use its definition, only its properties as proved in [Sar01]. Writing $p_a^*[\Phi] = \sup\{p \mid \gamma(p) < \infty\}$, the discriminant is $\Delta_a[\Phi] = \sup\{\gamma(p) \mid p < p_a^*[\Phi]\}$. By [Sar01, Proposition 3], the function $\gamma(p)$ is continuous and strictly increasing on $(-\infty, p_a^*[\Phi]]$.

We show that $\Delta_a[\Phi] > 0$ iff (2.10) holds. By [Sar01, (2),(4),(6)], we have

$$\Delta_a[\Phi] = \gamma(p_a^*[\Phi]), \tag{7.7}$$

$$p_a^*[\Phi] = -\overline{\lim}\, \frac{1}{n} \log Z_n^*(\Phi, a), \tag{7.8}$$

$$P_G(\Phi) = \begin{cases} -p(\Phi) & \Delta_a[\Phi] \ge 0, \\ -p_a^*[\Phi] & \Delta_a[\Phi] < 0, \end{cases} \tag{7.9}$$

where $p(\Phi)$ is the unique solution of $\gamma(p) = 0$, which exists iff $\Delta_a[\Phi] \ge 0$.

Now if $\Delta_a[\Phi] < 0$, then (7.8)–(7.9) show that $P_G(\Phi) = \overline{\lim}\, \frac{1}{n} Z_n^*(\Phi, a)$, so that (2.10) fails. It remains to consider the case when $\Delta_a[\Phi] \ge 0$, so the two sides of (2.10) are given by $-p_a^*[\Phi]$ and $-p(\Phi)$, where $\gamma(p(\Phi)) = 0$ and $\gamma(p_a^*[\Phi]) = \Delta_a[\Phi]$. Since $\gamma(p)$ is strictly increasing in p, it follows that $\Delta_a[\Phi] > 0$ iff $p(\Phi) < p_a^*[\Phi]$, which is equivalent to (2.10).

Proof of Proposition 3.16. Let $(\tilde{X}, \tilde{\sigma})$ be a shift factor of (X, σ). We prove that $h_{\mathrm{spec}}^{\perp}(\tilde{X}) \le h_{\mathrm{spec}}^{\perp}(X)$ by showing that if $\mathcal{C}^{\pm} \subset \mathcal{L}$ satisfy (1.2) and $[\mathbf{I}^*]$, then there are $\tilde{\mathcal{C}}^{\pm} \subset \tilde{\mathcal{L}}$ satisfying the same conditions and with the property that $h(\tilde{\mathcal{C}}^- \cup \tilde{\mathcal{C}}^+) \le h(\mathcal{C}^- \cup \mathcal{C}^+)$.

We follow the proof of [CT12, Proposition 2.2]: given two shifts $X, \tilde{X}$ on finite alphabets $A, \tilde{A}$ with a factor map $\pi \colon X \to \tilde{X}$, there is some $m \in \mathbb{N}$ and $\theta \colon \mathcal{L}_{2m+1}(X) \to \tilde{A}$ such that $\pi(x)_n = \theta(x_{[n-m,n+m]})$ for every $x \in X$

and $n \in \mathbb{Z}$. Writing $\Theta\colon \mathcal{L}_{n+2m} \to \tilde{\mathcal{L}}_n$ for the map induced by θ, we consider $\mathcal{C}^\pm \subset \mathcal{L}$ satisfying (1.2) and $[\mathbf{I}^*]$, and put

$$\tilde{\mathcal{C}}^- = \Theta(\mathcal{C}^-), \qquad \tilde{\mathcal{C}}^+ = \Theta(\mathcal{C}^+).$$

Since $\#\tilde{\mathcal{C}}_n^- \leq \#\mathcal{C}_{n+2m}^-$, and similarly for $\tilde{\mathcal{C}}^+$, we get $h(\tilde{\mathcal{C}}^- \cup \tilde{\mathcal{C}}^+) \leq h(\mathcal{C}^- \cup \mathcal{C}^+)$.

To prove (1.2) for $\tilde{\mathcal{C}}^+$, observe that given $\tilde{w} \in \tilde{\mathcal{C}}^+$ and $1 \leq i \leq |w|$, there is $w \in \mathcal{C}^+$ such that $\tilde{w} = \Theta(w)$, and in particular $\tilde{w}_{[1,i]} = \Theta(w_{[1,i+2m]}) \in \Theta(\mathcal{C}^+) = \tilde{\mathcal{C}}^+$ since $\mathcal{C}^+$ satisfies (1.2). The proof for $\tilde{\mathcal{C}}^-$ is similar.

Finally, every $\mathcal{G}^M(\tilde{\mathcal{C}}^\pm)$ has specification in the sense of $[\mathbf{I}^*]$: given $\tilde{v} \in \mathcal{G}^M(\tilde{\mathcal{C}}^\pm)$, let $v \in \mathcal{L}$ be such that $\tilde{v} = \Theta(v)$. Then $\tilde{v}_{[1,i]} \notin \tilde{\mathcal{C}}^-$ for all $i > M$ implies that $v_{[1,j]} \notin \mathcal{C}^-$ for all $j > M + 2m$, and we similarly deduce that $v_{[i,|v|]} \notin \mathcal{C}^+$ for $i \leq |v| - M - 2m$, hence $v \in \mathcal{G}^{M+2m}(\mathcal{C}^\pm)$.

Now given $\tilde{v}, \tilde{w} \in \mathcal{G}^M(\tilde{\mathcal{C}}^\pm)$, there are $v, w \in \mathcal{G}^{M+2m}(\mathcal{C}^\pm)$ such that $\Theta(v) = \tilde{v}$ and $\Theta(w) = \tilde{w}$. By $[\mathbf{I}^*]$ for $\mathcal{C}^\pm$ there is $u \in \mathcal{L}$ such that $|u| \leq \tau(M + 2m)$ and $vuw \in \mathcal{L}$. Thus $\Theta(vuw) = \tilde{v} \cdot \Theta(v_{(|v|-2m,|v|]}uw_{[1,2m]}) \cdot \tilde{w} \in \tilde{\mathcal{L}}$, so $\mathcal{G}^M(\tilde{\mathcal{C}}^\pm)$ satisfies $[\mathbf{I}^*]$ with $\tilde{\tau}(M) = \tau(M + 2m) + 2m$.

Proof of Proposition 3.17. Let (X, σ) be a shift space with $\mathcal{G} \subset \mathcal{L}(X)$ satisfying $[\mathbf{I}]$ and $[\mathbf{E}^*]$. If $\mathcal{G}$ is periodic in the sense of Proposition 6.1, then by $[\mathbf{E}^*]$, $\mathcal{L}$ is periodic as well, so X is a single periodic orbit. Thus if X is not a single periodic orbit, Lemma 6.6 applies to give $h(\mathcal{G}) > 0$, hence $h(X) > 0$.

If $\tilde{X}$ is a shift factor of X with factor map $\Theta\colon \mathcal{L}_{n+2m} \to \tilde{\mathcal{L}}_n$, then taking $\tilde{\mathcal{G}} = \Theta(\mathcal{G})$ we see that $\tilde{\mathcal{G}}$ has $[\mathbf{I}]$; indeed, given any $\tilde{v}, \tilde{w} \in \tilde{\mathcal{G}}$ we take $v, w \in \mathcal{G}$ such that $\Theta(v) = \tilde{v}$ and $\Theta(w) = \tilde{w}$, then there is $u \in \mathcal{L}_{\leq \tau}$ such that $vuw \in \mathcal{G}$, and we see that $\Theta(vuw) = \tilde{v}\tilde{u}\tilde{w} \in \tilde{\mathcal{G}}$, where $|\tilde{u}| \leq \tau + 2m$, hence $\tilde{\mathcal{G}}$ has $[\mathbf{I}]$. Moreover, $\tilde{\mathcal{L}}$ has $[\mathbf{E}^*]$, since any $\tilde{w} \in \tilde{\mathcal{L}}$ has $\tilde{w} = \Theta(w)$ for some $w \in \mathcal{L}$, and by $[\mathbf{E}^*]$ for $\mathcal{L}$ there are $u, v \in \mathcal{L}$ such that $uwv \in \mathcal{G}$, hence $\tilde{u}\tilde{w}\tilde{v} := \Theta(uwv) \in \tilde{\mathcal{G}}$.

We have shown that $[\mathbf{I}]$ and $[\mathbf{E}^*]$ pass to factors; it remains only to show that if $\gcd\{k \mid \mathrm{Per}_k(X) \neq \emptyset\} = 1$, then $\tilde{X}$ is not a single non-trivial periodic orbit. For this it suffices to observe that if $x \in \mathrm{Per}_k(X)$, then $\sigma^k(x) = x$ and hence $\tilde{\sigma}^k(\Theta(x)) = \Theta(\sigma^k(x)) = \Theta(x)$, hence $\Theta(x)$ is periodic with period a factor of k. If $\tilde{X}$ is a single periodic orbit with least period p, then p divides k whenever $\mathrm{Per}_k(X) \neq \emptyset$, and hence $p = 1$.

Proof of claims in Remark 3.3. Recall the example described in Remark 3.3; fix $k \geq 4$ and let X be the SFT on $A = \{1, \dots, k\}$ such that the allowed transitions are $a \to a+1 \pmod k$ and $a \to a+2 \pmod k$. Suppose $\mathcal{F} \subset \mathcal{L}(X)$ satisfies $[\mathbf{I}_0]$. Let $B = \{w_1 \mid w \in \mathcal{F}\}$ and $C = \{w_{|w|} \mid w \in \mathcal{F}\}$. Then by $[\mathbf{I}_0]$ we have $c \to b$ for every $c \in C$ and $b \in B$. Since each $a \in A$ has exactly two followers (two choices of b such that $a \to b$) and no two choices of a have the same set of two followers, one of B, C must be a singleton, call it $\{a\}$. Then every word in $\mathcal{F}$ either starts or ends with a, and we see that for any choice of $\mathcal{E}^p, \mathcal{E}^s$, we have $\mathcal{E} \supset \{w \in \mathcal{L} \mid w_j \neq a \text{ for all } 1 \leq j \leq |w|\} =: \mathcal{D}$.

Because every state has two followers we see that $h(X) = \log 2$ (there are always two choices for the next symbol, so $\#\mathcal{L}_n = k2^{n-1}$). On the other hand, we can estimate $h(\mathcal{D})$ from below as follows: given $n \in \mathbb{N}$ and $u \in \{1,2\}^n$, define $\pi(u) \in \mathcal{D}_{n+1}$ by $\pi(u)_1 = 1$ and

$$\pi(u)_{i+1} = \begin{cases} a-1 & \text{if } \pi(u)_i = a-2, \\ a+1 & \text{if } \pi(u)_i = a-1, \\ \pi(u)_i + u_i & \text{otherwise,} \end{cases}$$

where we work mod k. Given $u, v \in \{1,2\}^n$, we have $\pi(u) = \pi(v)$ if and only if $u_i = v_i$ for all i such that $\pi(u)_i \notin \{a-2, a-1\}$, and since every index interval of length $k/2$ contains at most 2 values of i with $\pi(u)_i \in \{a-2, a-1\}$, we see that every $w \in \mathcal{D}_{n+1}$ has $\#\pi^{-1}(w) \le 2^{2\frac{n}{k/2}}$, so $\#\mathcal{D}_{n+1} \ge 2^n 2^{-4n/k}$, and we get $h(\mathcal{E}) \ge h(\mathcal{D}) \ge (1 - \frac{4}{k}) \log 2$. This shows that for every choice of $\mathcal{E}^p, \mathcal{F}, \mathcal{E}^s$, we have $h(\mathcal{E}) > 0$,[29] and that the entropy gap between $h(\mathcal{E})$ and $h(X)$ can be forced to be arbitrarily small by taking d large.

References

[ADLP16] José F. Alves, C. L. Dias, S. Luzzatto, and V. Pinheiro, *SRB measures for partially hyperbolic systems whose central direction is weakly expanding*, to appear in Journal of the European Math. Society, arXiv:1403.2937, 2016.

[AHS14] Artur Avila, Pascal Hubert, and Alexandra Skripchenko, *Diffusion for chaotic plane sections of 3-periodic surfaces*, Preprint. `arXiv:1412.7913`, 2014.

[AL15] José F. Alves and Xin Li, *Gibbs–Markov–Young structures with (stretched) exponential tail for partially hyperbolic attractors*, Advances in Mathematics **279** (2015), 405 – 437.

[AP10] José F. Alves and Vilton Pinheiro, *Gibbs-Markov structures and limit laws for partially hyperbolic attractors with mostly expanding central direction*, Adv. Math. **223** (2010), no. 5, 1706–1730. MR 2592507 (2011e:37068)

[BCLS15] Martin Bridgeman, Richard Canary, François Labourie, and Andres Sambarino, *The pressure metric for Anosov representations*, Geom. Funct. Anal. **25** (2015), no. 4, 1089–1179. MR 3385630

[Ber88] Anne Bertrand, *Specification, synchronisation, average length*, Coding theory and applications (Cachan, 1986), Lecture Notes in Comput. Sci., vol. 311, Springer, Berlin, 1988, pp. 86–95. MR 960710 (89i:94022)

[BG11] A. I. Bufetov and B. M. Gurevich, *Existence and uniqueness of a measure with maximal entropy for the Teichmüller flow on the moduli space of abelian differentials*, Mat. Sb. **202** (2011), no. 7, 3–42. MR 2857792 (2012h:37068)

[BH86] F. Blanchard and G. Hansel, *Systèmes codés*, Theoret. Comput. Sci. **44** (1986), no. 1, 17–49. MR 858689 (88m:68029)

[BL13] Henk Bruin and Renaud Leplaideur, *Renormalization, thermodynamic formalism and quasi-crystals in subshifts*, Comm. Math. Phys. **321** (2013), no. 1, 209–247. MR 3089670

[Bow72] Rufus Bowen, *The equidistribution of closed geodesics*, Amer. J. Math. **94** (1972), 413–423. MR 0315742 (47 #4291)

[Bow74] ———, *Some systems with unique equilibrium states*, Math. Systems Theory **8** (1974), no. 3, 193–202. MR 0399413 (53 #3257)

[29] When $k = 4$ we have $1 - \frac{4}{k} = 0$, but a more careful inspection shows that $h(\mathcal{D}) > 0$.

[Bow75] ———, *Equilibrium states and the ergodic theory of Anosov diffeomorphisms*, Lecture Notes in Mathematics, Vol. 470, Springer-Verlag, Berlin-New York, 1975. MR 0442989 (56 #1364)

[Bow78] ———, *On Axiom A diffeomorphisms*, American Mathematical Society, Providence, R.I., 1978, Regional Conference Series in Mathematics, No. 35. MR 0482842

[Boy08] Mike Boyle, *Open problems in symbolic dynamics*, Geometric and probabilistic structures in dynamics, Contemp. Math., vol. 469, Amer. Math. Soc., Providence, RI, 2008, pp. 69–118. MR 2478466 (2010h:37023)

[BS03] Jérôme Buzzi and Omri Sarig, *Uniqueness of equilibrium measures for countable Markov shifts and multidimensional piecewise expanding maps*, Ergodic Theory Dynam. Systems **23** (2003), no. 5, 1383–1400. MR 2018604 (2004k:37053)

[BSS02] Luis Barreira, Benoît Saussol, and Jörg Schmeling, *Higher-dimensional multifractal analysis*, J. Math. Pures Appl. (9) **81** (2002), no. 1, 67–91. MR 1994883 (2004g:37038)

[BT09] Henk Bruin and Mike Todd, *Equilibrium states for interval maps: the potential* $-t\log|Df|$, Ann. Sci. Éc. Norm. Supér. (4) **42** (2009), no. 4, 559–600. MR 2568876 (2011c:37079)

[Buz97] Jérôme Buzzi, *Intrinsic ergodicity of affine maps in* $[0,1]^d$, Monatsh. Math. **124** (1997), no. 2, 97–118. MR 1462857 (98g:58098)

[Buz99] ———, *Markov extensions for multi-dimensional dynamical systems*, Israel J. Math. **112** (1999), 357–380. MR 1714974 (2000m:37008)

[Buz04] ———, *Entropy of equilibrium measures of continuous piecewise monotonic maps*, Stoch. Dyn. **4** (2004), no. 1, 84–94. MR 2069369 (2005h:37071)

[Buz05] ———, *Subshifts of quasi-finite type*, Invent. Math. **159** (2005), no. 2, 369–406. MR 2116278 (2005i:37013)

[CFT15] Vaughn Climenhaga, Todd Fisher, and Daniel J. Thompson, *Unique equilibrium states for the robustly transitive diffeomorphisms of Mañé and Bonatti–Viana*, Preprint, arXiv:1505.06371, 2015.

[Con] Scott Conrad, *A coded shift with a Hölder potential that is not hyperbolic*, preprint.

[CP16] Vaughn Climenhaga and Ronnie Pavlov, *One-sided almost specification and intrinsic ergodicity*, Preprint, 2016.

[CS09] Van Cyr and Omri Sarig, *Spectral gap and transience for Ruelle operators on countable Markov shifts*, Comm. Math. Phys. **292** (2009), no. 3, 637–666. MR 2551790 (2011f:37041)

[CT12] Vaughn Climenhaga and Daniel J. Thompson, *Intrinsic ergodicity beyond specification: β-shifts, S-gap shifts, and their factors*, Israel J. Math. **192** (2012), no. 2, 785–817. MR 3009742

[CT13] ———, *Equilibrium states beyond specification and the Bowen property*, J. Lond. Math. Soc. (2) **87** (2013), no. 2, 401–427. MR 3046278

[CT14] ———, *Intrinsic ergodicity via obstruction entropies*, Ergodic Theory Dynam. Systems **34** (2014), no. 6, 1816–1831. MR 3272773

[CT15] Vaughn Climenhaga and Daniel J. Thompson, *Unique equilibrium states for flows and homeomorphisms with non-uniform structure*, Preprint, arXiv:1505.03803, 2015.

[CTY] Vaughn Climenhaga, Daniel J. Thompson, and Kenichiro Yamamoto, *Large deviations for systems with non-uniform structure*, Trans. Amer. Math. Soc., 32 pages, to appear. `arXiv:1304.5497`.

[FF92] Doris Fiebig and Ulf-Rainer Fiebig, *Covers for coded systems*, Symbolic dynamics and its applications (New Haven, CT, 1991), Contemp. Math., vol.

135, Amer. Math. Soc., Providence, RI, 1992, pp. 139–179. MR 1185086 (93m:54068)

[Hof78] Franz Hofbauer, *β-shifts have unique maximal measure*, Monatsh. Math. **85** (1978), no. 3, 189–198. MR 0492180 (58 #11326)

[Hof79] ———, *On intrinsic ergodicity of piecewise monotonic transformations with positive entropy*, Israel J. Math. **34** (1979), no. 3, 213–237 (1980). MR 570882 (82c:28039a)

[Hof81] ———, *On intrinsic ergodicity of piecewise monotonic transformations with positive entropy. II*, Israel J. Math. **38** (1981), no. 1-2, 107–115. MR 599481 (82c:28039b)

[IRRL12] Irene Inoquio-Renteria and Juan Rivera-Letelier, *A characterization of hyperbolic potentials of rational maps*, Bull. Braz. Math. Soc. (N.S.) **43** (2012), no. 1, 99–127. MR 2909925

[IT10] Godofredo Iommi and Mike Todd, *Natural equilibrium states for multimodal maps*, Comm. Math. Phys. **300** (2010), no. 1, 65–94. MR 2725183 (2011m:37048)

[Kat82] A. Katok, *Entropy and closed geodesics*, Ergodic Theory Dynam. Systems **2** (1982), no. 3-4, 339–365 (1983). MR 721728 (85b:53047)

[Kel89] Gerhard Keller, *Lifting measures to Markov extensions*, Monatsh. Math. **108** (1989), no. 2-3, 183–200. MR 1026617 (91b:28011)

[Kif90] Yuri Kifer, *Large deviations in dynamical systems and stochastic processes*, Trans. Amer. Math. Soc. **321** (1990), no. 2, 505–524. MR 1025756 (91e:60091)

[Kni98] Gerhard Knieper, *The uniqueness of the measure of maximal entropy for geodesic flows on rank* 1 *manifolds*, Ann. of Math. (2) **148** (1998), no. 1, 291–314. MR 1652924 (2000b:37016)

[Kwa00] Jaroslaw Kwapisz, *Cocyclic subshifts*, Math. Z. **234** (2000), no. 2, 255–290. MR 1765882 (2001j:37025)

[Kwa04] ———, *Transfer operator, topological entropy and maximal measure for cocyclic subshifts*, Ergodic Theory Dynam. Systems **24** (2004), no. 4, 1173–1197. MR 2085908 (2005d:37018)

[Liv96] Carlangelo Liverani, *Central limit theorem for deterministic systems*, International Conference on Dynamical Systems (Montevideo, 1995), Pitman Res. Notes Math. Ser., vol. 362, Longman, Harlow, 1996, pp. 56–75. MR 1460797 (98k:28025)

[LM95] Douglas Lind and Brian Marcus, *An introduction to symbolic dynamics and coding*, Cambridge University Press, Cambridge, 1995. MR 1369092 (97a:58050)

[LRL14] Huaibin Li and Juan Rivera-Letelier, *Equilibrium states of weakly hyperbolic one-dimensional maps for Hölder potentials*, Comm. Math. Phys. **328** (2014), no. 1, 397–419. MR 3196990

[McM08] Curtis T. McMullen, *Thermodynamics, dimension and the Weil-Petersson metric*, Invent. Math. **173** (2008), no. 2, 365–425. MR 2415311 (2010c:37101)

[MN05] Ian Melbourne and Matthew Nicol, *Almost sure invariance principle for nonuniformly hyperbolic systems*, Comm. Math. Phys. **260** (2005), no. 1, 131–146. MR 2175992 (2006h:37047)

[NP66] D. Newton and W. Parry, *On a factor automorphism of a normal dynamical system*, Ann. Math. Statist **37** (1966), 1528–1533. MR 0206209 (34 #6028)

[Pet89] Karl Petersen, *Ergodic theory*, Cambridge Studies in Advanced Mathematics, vol. 2, Cambridge University Press, Cambridge, 1989, Corrected reprint of the 1983 original. MR 1073173 (92c:28010)

[PP90] William Parry and Mark Pollicott, *Zeta functions and the periodic orbit structure of hyperbolic dynamics*, Astérisque (1990), no. 187-188, 268. MR 1085356 (92f:58141)

[PS08] Yakov Pesin and Samuel Senti, *Equilibrium measures for maps with inducing schemes*, J. Mod. Dyn. **2** (2008), no. 3, 397–430. MR 2417478 (2009k:37072)

[PSZ08] Ya. B. Pesin, S. Senti, and K. Zhang, *Lifting measures to inducing schemes*, Ergodic Theory Dynam. Systems **28** (2008), no. 2, 553–574. MR 2408392 (2009f:37004)

[PSZ14] Yakov Pesin, Samuel Senti, and Ke Zhang, *Thermodynamics of towers of hyperbolic type*, to appear in Trans. Amer. Math. Soc., arXiv:1403.2989, 2014.

[PZ07] Yakov Pesin and Ke Zhang, *Thermodynamics of inducing schemes and liftability of measures*, Partially hyperbolic dynamics, laminations, and Teichmüller flow, Fields Inst. Commun., vol. 51, Amer. Math. Soc., Providence, RI, 2007, pp. 289–305. MR 2388701 (2009a:37062)

[Rue76] David Ruelle, *A measure associated with axiom-A attractors*, Amer. J. Math. **98** (1976), no. 3, 619–654. MR 0415683 (54 #3763)

[Rue92] ———, *Thermodynamic formalism for maps satisfying positive expansiveness and specification*, Nonlinearity **5** (1992), no. 6, 1223–1236. MR 1192516 (94a:58115)

[Sar99] Omri M. Sarig, *Thermodynamic formalism for countable Markov shifts*, Ergodic Theory Dynam. Systems **19** (1999), no. 6, 1565–1593. MR 1738951 (2000m:37009)

[Sar01] ———, *Phase transitions for countable Markov shifts*, Comm. Math. Phys. **217** (2001), no. 3, 555–577. MR 1822107 (2002b:37040)

[Sar11] ———, *Bernoulli equilibrium states for surface diffeomorphisms*, J. Mod. Dyn. **5** (2011), no. 3, 593–608. MR 2854097 (2012k:37071)

[Sar13] ———, *Symbolic dynamics for surface diffeomorphisms with positive entropy*, J. Amer. Math. Soc. **26** (2013), no. 2, 341–426. MR 3011417

[Sar15] ———, *Thermodynamic formalism for countable Markov shifts*, Hyperbolic dynamics, fluctuations and large deviations, Proc. Sympos. Pure Math., vol. 89, Amer. Math. Soc., Providence, RI, 2015, pp. 81–117. MR 3309096

[Sch97] Jörg Schmeling, *Symbolic dynamics for β-shifts and self-normal numbers*, Ergodic Theory Dynam. Systems **17** (1997), no. 3, 675–694. MR 1452189 (98c:11080)

[Tho06] Klaus Thomsen, *On the ergodic theory of synchronized systems*, Ergodic Theory Dynam. Systems **26** (2006), no. 4, 1235–1256. MR 2247640 (2009a:37011)

[Wal78] Peter Walters, *Equilibrium states for β-transformations and related transformations*, Math. Z. **159** (1978), no. 1, 65–88. MR 0466492 (57 #6370)

[Wal82] ———, *An introduction to ergodic theory*, Graduate Texts in Mathematics, vol. 79, Springer-Verlag, New York-Berlin, 1982. MR 648108 (84e:28017)

[You98] Lai-Sang Young, *Statistical properties of dynamical systems with some hyperbolicity*, Ann. of Math. (2) **147** (1998), no. 3, 585–650. MR 1637655 (99h:58140)

[You99] ———, *Recurrence times and rates of mixing*, Israel J. Math. **110** (1999), 153–188. MR 1750438 (2001j:37062)

Department of Mathematics, University of Houston, Houston, Texas 77204
E-mail address: `climenha@math.uh.edu`